

AD-A248 772

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM - 1991

GRADUATE STUDENT RESEARCH PROGRAM
(GSRP) REPORTS

VOLUME 9

WRIGHT LABORATORY



RESEARCH & DEVELOPMENT LABORATORIES

5800 UPLANDER WAY

CULVER CITY, CA 90230-6603

SUBMITTED TO:

LT. COL. CLAUDE CAVENDER
PROGRAM MANAGER

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE

WASHINGTON, D.C.

DECEMBER 1991

AFSC 92 0175

SCIENTIFIC RESEARCH (AFSC)
UNCLASSIFIED TO DTIC
This report has been reviewed and is
being made public release IAW AFR 190-12
DTIC is unlimited.
STINGO Program Manager

Approved for public release,
distribution unlimited

92-09034



92 4 08 011

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, completing and reviewing the collection of information, sending comments regarding this burden estimate or any other aspect of this collection of information, including the burden of Washington Headquarters Services, Directorate for Information Operations and Reports, 215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 9 January 1992	3. REPORT TYPE AND DATES COVERED 30 Sep 90-30 Sep 91	
4. TITLE AND SUBTITLE 1991 Graduate Student Research Program (GSRP) Volumes 6-9 Vol. 9			5. FUNDING NUMBERS F49620-90-C-0076	
5. AUTHOR(S) Mr Gary Moore				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research Development Laboratories (RDL) 5800 Uplander Way Culver City CA 90230-6608 AFOSR-TR-			8. PERFORMING ORGANIZATION REPORT NUMBER 92 0176	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NI Bldg 410 Bolling AFB DC 20332-644P Lt Col V. Claude Cavender			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT UNLIMITED			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This program was started in 1982 as an adjunct to the SFRP. Its objectives are to permit graduate students to participate in research under the direction of a faculty member at an Air Force laboratory; stimulate professional association among graduate students, their supervising professors, and professional peers in the Air Force; to further research objectives of the Air Force; and to expose graduate students to potential thesis topics in areas of interest to the Air Force. During the summer of 1991 142 graduate students performed research for 10 weeks at Air Force laboratories. Their reports were submitted to RDL and consolidated into this annual report.				
14. SUBJECT TERMS			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1991
GRADUATE STUDENT RESEARCH PROGRAM (GSRP) REPORTS

VOLUME 9

WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Lt. Col. Claude Cavender

Program Manager, RDL
Claude Baum

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1991



Decision For	
AFOSR	CRA&I <input checked="" type="checkbox"/>
AFOSR	TAB <input type="checkbox"/>
AFOSR	Announced <input type="checkbox"/>
Justification	
By	
Distribution	
Availability Codes	
Dist	Availability or Special
A-1	

PREFACE

Reports in this document are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

This document is one of a set of 13 volumes describing the 1991 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME

TITLE

1 Program Management Report

Summer Faculty Research Program (SFRP) Reports

2 Armstrong Laboratory, Wilford Hall Medical Center

3 Phillips Laboratory, Civil Engineering Laboratory

4 Rome Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory

5 Wright Laboratory

Graduate Student Research Program (GSRP) Reports

6 Armstrong Laboratory, Wilford Hall Medical Center

7 Phillips Laboratory, Civil Engineering Laboratory

8 Rome Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory

9 Wright Laboratory

High School Apprenticeship Program (HSAP) Reports

10 Armstrong Laboratory

11 Phillips Laboratory, Civil Engineering Laboratory

12 Rome Laboratory, Arnold Engineering Development Center

13 Wright Laboratory

1991 GRADUATE STUDENT RESEARCH REPORTS

Wright Laboratory

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
Aero Propulsion & Power Laboratory (PROP)		
1	The Use of Thermographic Phosphors for Surface Temperature Measurement	David Bizzak
2	Experimental and Analytical Investigation of Effects of Noncondensable Gases on On-Axis Rotating Heat Pipes	Bryan Martin
3	A Two-Wire Thermocouple Technique for Dynamic Temperature Measurement in Reacting Flows	Thomas Scattergood
4	Investigation of the Combustion Characteristics of Swirled Injectors in a Confined Coannular System with a Sudden Expansion	David Warren
5	Not Available at this Time	Mark Williams
Armament Laboratory (ATL)		
6	Two Dimensional Simulation of Railgun Plasma Armatures	George Boynton
7	Setup of Multispectral Radiometer System	Laurence Flath
8	Laser Polarimetry	Randy Gove
9	Material Characterization and Evaluation for Penetrator Applications	Steven Hatfield
10	Two-Dimensional Waveguide Simulation	Christopher Hussell
11	Materials for Infrared Modulator Technology	Annette Marsh
12	Material Characterization and Evaluation for Penetrator Applications	Jessica Mayes
13	Full Reynolds-Averaged Navier-Stokes Simulation of Turbulent Transonic Flow Past a Wrap-Around Fin Projectile	Juan Ordonez
14	Implementation of a System for Evaluating the Modulation Transfer Function of CCDs Using Laser Speckle	Martin Sensiper
Avionics Laboratory (AVION)		
15	A Note on Prony's Method	Wendy Barkman
16	Applying Cascade Correlation to the Extrapolation of Chaotic Time Series	David Ensley
17	Computer Aided Analysis of LPI Signal Detectability	Scott Francis

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
<u>Wright Laboratory (cont.)</u>		
18	Minimum Resolvable Temperature Measurements of the Mitsubishi Thermal Imager IR 5210C	Darron Lockett
19	Performance Evaluation of Rule Grouping Algorithm Running on the Activation Framework Architecture	Bryant Poole
Electronic Technology Laboratory (ETL)		
20	High Temperature OHMIC Contacts for GaAs	David Moss
Flight Dynamic Laboratory (FDL)		
21	Acoustic Emission Behavior and Damage Accumulation Monitoring of Cord-Rubber Composite Materials	Pat Hippo
22	De ply of Laminated Panels with Perforation due to Impact	John Lair
23	Actuator Location and Optimal Control Design for Flexible Structures	John Mitchell
24	Finite Element Non-Linear Multimode Random Response of Complex Panels	Russell Morris
25	Evaluation of the Analytical Design Package (ADP) for Frameless Transparency Program	David Russell
26	Experimental Investigation of the Influence of Constrained-Layer Damping Treatment on Parametric and Autoparametric Resonances in Nonlinear Structural Systems	Joseph Schudt
27	Application of Computational Fluid Dynamics for Optimal Structural Design of Flight Vehicles	Anne Stephenson
28	Ballistic Impact Test of Graphite/Epoxy (Composite Material)	Aziz Tadayon
29	Hypersonic Vehicle Control Strategies: Preliminary Considerations	Phuong Vu
Materials Laboratory (MAT)		
30	A Study on the Effects of Temperature and Gas Composition on the Microstructure/Morphology of CVD SiC in the MTS/H ₂ System	Peter Brown
31	Thermal Degradation of Perfluoropolyalkyether Fluids in the Presence of Various Metals	David Drake
32	Numerical Simulation of Photoreflectance from Two-Dimensional Electron Gas in a GaAs/AiGaAs Heterojunction	James Engholm
33	Pulsed Laser Deposition and Near Optimal Forging Accomplishments	Douglas Moore

**Report
Number**

Report Title

Author

Wright Laboratory (cont.)

- | | | |
|----|---|-------------------|
| 34 | Decomposition Pathways for Model Fluorinated Ether Lubricants on Iron, Chromium and Nickel Surfaces | Mary Napier |
| 35 | Ultrasonic Beam Propagation: Diffractionless Beams and Beams in Anisotropic Media | Mark Preischel |
| 36 | Polymer Conductivity and Optics | Joanne Promislow |
| 37 | Mathematical Modeling of Carbon/Carbon Processing | Jennifer Reid |
| 38 | Automated Inspection Code Generation | Steven Ruegsegger |
| 39 | Processing of Monolithic Ti_3Si_2 | Andrew Thom |

THE USE OF THERMOGRAPHIC PHOSPHORS FOR SURFACE TEMPERATURE MEASUREMENT

David J. Bizzak
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

ABSTRACT

The efficacy of a nonintrusive, optical technique for surface temperature measurement based on the temperature sensitivity of the fluorescent emission spectrum of rare-earth ion doped phosphors (i.e., thermographic phosphors) is demonstrated in an experiment with a heated circular cylinder in cross flow. In this experiment the cylinder is heated by a constant surface heat flux, which is provided by applying a voltage across the foil surface of the cylinder. To allow surface temperature measurement, a thin coating of a thermographic phosphor is applied to the foil. The phosphor fluoresces when excited by ultraviolet (UV) illumination provided by a mercury vapor lamp. The fluorescent emission at two different wavelengths are captured by a charge coupled device (CCD) video camera. The analog output of the CCD camera is digitized and stored using PC-based hardware. These thermographic phosphor image data provide the means to calculate discrete circumferential surface temperatures, while thermocouples attached to the inner surface of the foil cylinder provide collaborating measurements. Comparison of circumferential surface temperatures obtained using the thermographic phosphor with those recorded from the surface-mounted thermocouples indicate that the accuracy of the present technique is ± 2 °F, but future technique improvements promise to significantly increase the accuracy of thermographic phosphor temperature measurements.

INTRODUCTION

Although the thermocouple is the most common temperature measurement device used in heat transfer research, their usage to evaluate the heat transfer characteristics of high-performance gas turbines presents several technical difficulties. First, a proper examination of the heat transfer and thermal stresses in key components such as the turbine blades requires the use of many thermocouples. The process of mounting numerous thermocouples on or within a modern turbine blade is complicated by the presence of

intricate cooling passages within the blade. Great care must be exercised to ensure that the thermocouple leads do not affect the flow within these coolant passages. An additional concern in the instrumentation of rotating turbine components is the use of slip ring assemblies, which permit intermittent contact between rotating thermocouple leads and stationary data acquisition system leads. The movement of these contacts introduces noise in the output signal. More importantly, however, such assemblies are not suitable for use at the very high rotation rates associated with modern, high-performance aircraft turbines. Consequently, it is not possible to obtain temperature data at normal operating conditions. Therefore, nominal heat transfer characteristics must be inferred from lower speed tests.

To overcome these limitations, WL/POTC has investigated the feasibility of developing a nonintrusive, optical temperature and heat flux measurement system based on the temperature-sensitive properties of rare-earth ion doped phosphors. The fluorescent emission spectrum of such phosphors is characterized by unique emission lines, which correspond to different system energy level. Because the energy level population distribution is known to be a function of temperature, the ratio of the intensity of an upper energy level emission line to that of a lower energy level can be accurately correlated with temperature. In this study europium-doped lanthanum oxysulfide ($\text{La}_2\text{O}_2\text{S}:\text{Eu}$) is calibrated for temperatures from 80 °F to 130 °F by ratioing the intensity of the fluorescent emissions at 510 nm and 620 nm.

EXPERIMENTAL APPARATUS AND PROCEDURE

The key component of the experimental setup is the test cylinder. It is constructed of a 6-inch (152 mm) by 11-inch (279 mm) sheet of thin stainless steel conducting foil wrapped around 3-inch (76 mm) diameter brass disks located near the top and bottom of the foil. The outer surface of the foil is coated with a thin film of the $\text{La}_2\text{O}_2\text{S}:\text{Eu}$ phosphor, while the midplane of the inner surface is instrumented with thermocouples placed at 30° increments, beginning at the leading edge stagnation point. An electrically-conductive silver epoxy is used to bond the foil to the brass disks. By applying a constant voltage between the two disks, a constant surface heat flux can be produced. To prevent the formation of natural convective circulation cells within the cylinder that might alter local surface heat transfer during the experiment, the interior is filled with an insulating foam.

The test cylinder is shown schematically in Figure 1, along with the optical surface temperature measurement equipment. The components of this system serve two primary functions—phosphor excitation and image acquisition and processing. The phosphor

excitation is provided by the ultraviolet (UV) output of a mercury vapor lamp, while image acquisition is accomplished using a CCD video camera and a PC-based image acquisition card. The arrangement and use of these components is more fully described below.

A 100-watt mercury vapor lamp provides the UV excitation necessary to cause the $\text{La}_2\text{O}_2\text{S:Eu}$ to fluoresce. Since the plexiglas walls of the wind tunnel are effective absorbers of UV radiation, quartz viewing windows are installed to permit illumination of the test cylinder. In addition, an UV filter is used to eliminate that portion of the lamp output within the visible spectrum. This is necessary to minimize the noise associated with the inclusion of reflected background light when evaluating the intensity of the fluorescent emission.

The remaining system components provide the means to acquire and store images. The wavelength and bandwidth of the fluorescent emission captured by the video equipment is selected by positioning a narrow bandpass filter in front of the camera lens. For the $\text{La}_2\text{O}_2\text{S:Eu}$ phosphor, a 510 nm and a 620 nm bandpass filter with a bandwidth of 10 nm are used. A lens holder with a slide mechanism permits precise positioning of the bandpass filters. A 3-CCD Sony video camera equipped with a macro lens is used to provide an image of the illuminated portion of the cylinder. The analog video signal is digitized into a 512 by 512 pixel array by a Data Translation image acquisition card. Each pixel is assigned an 8-bit intensity value, which equates to 256 gray scale levels. While the phosphor fluorescent emission is quite intense when viewed by the naked eye, only a small fraction of the entire spectrum is analyzed. As a result, an 18 dB electrical gain is required to obtain images of sufficient intensity. Even with the excellent signal-to-noise ratio of the camera (67 dB), amplification will introduce signal noise. To minimize the effect of this noise, 64 video frames are averaged for each image acquired. The numerator (510 nm) and denominator (620 nm) images are acquired sequentially in real time and stored in binary files for processing after completion of the experiment.

Prior to conducting the experiment, selected regions of the phosphor surface are calibrated. The first step of this procedure is the focusing and alignment of the UV source. Though the output of a vapor lamp is nonuniform, intensity peaks can be smoothed by expanding the output beam. Beam expansion, however, decreases the illumination intensity per unit area. As UV intensity is decreased, a threshold is reached at which the intensity level of the phosphor fluorescent emission is insufficient to be reliably detected by the camera. Due to the relatively low power of the mercury vapor lamp used in this study and the requirement that the lamp remain outside the wind tunnel, no more than 45% of the

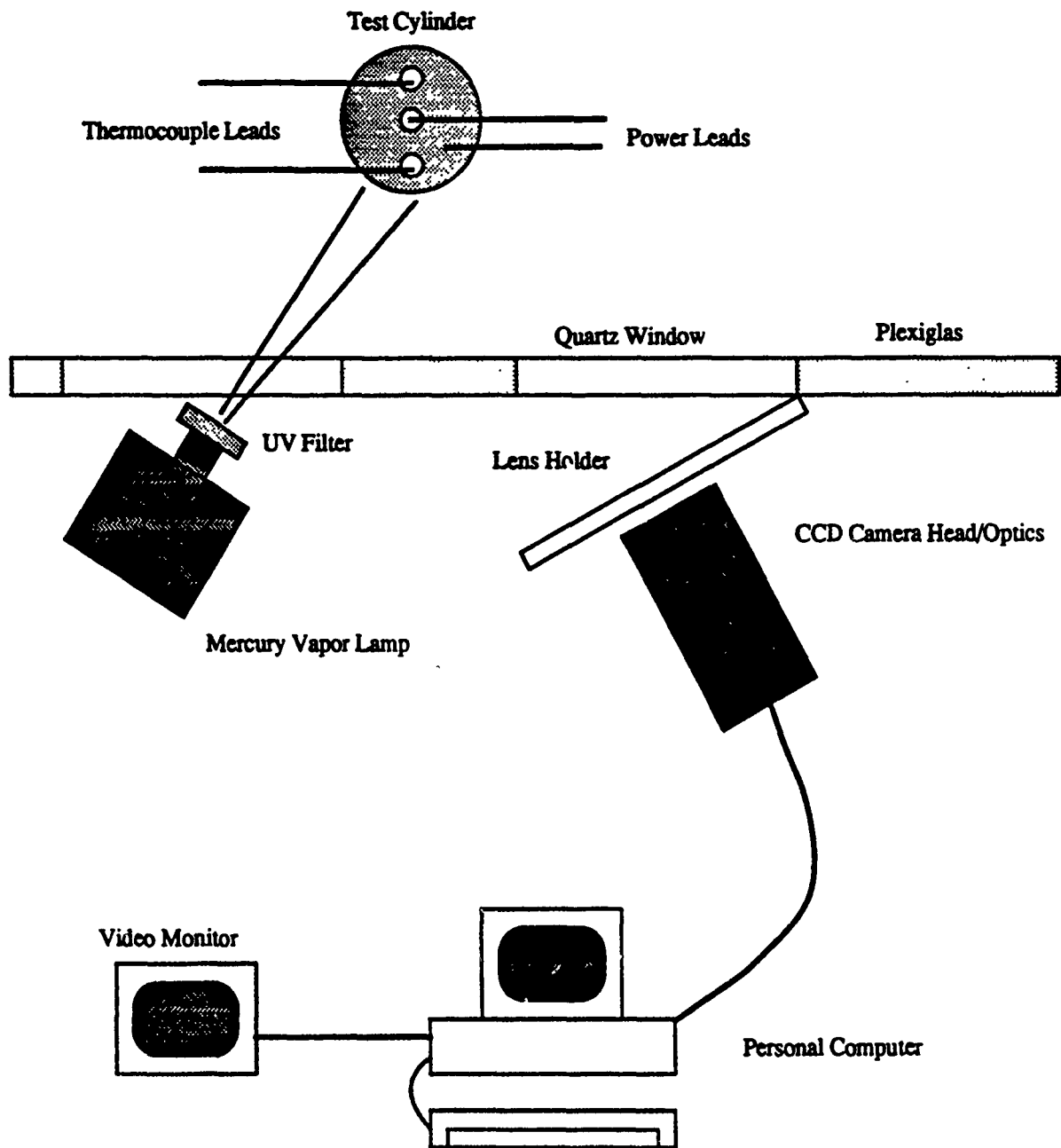


Figure 1 Experimental Setup

cylinder's surface can be effectively excited. After proper UV illumination is obtained, the regions of interest to be calibrated are selected. Using the surface alignment marks that are spaced at 15-degree intervals around the cylinder, it is possible to accurately define

rectangles centered at 7.5-degree increments on the lighted portion of the cylinder. Each rectangular region is 11 pixels wide by 25 pixels high. The width of each of these regions corresponds to approximately 1 degree of radial arc. Numerator and denominator images are captured for several temperatures between 80 °F and 130 °F. In this calibration, surface temperature is varied by changing the applied voltage across the foil surface. With the cylinder in still air, an equilibrium surface temperature at any voltage is reached when natural convective heat transfer equals the surface heat flux. This temperature is measured using the surface-mounted thermocouples, whose accuracy has been verified to be ± 0.2 °F over the temperature range of interest. The captured image data are processed to obtain a calibration curve for each rectangular region, relating emission intensity ratio to temperature.

After the illuminated portion of the cylinder is calibrated, tests at a Reynolds number of approximately 2.0×10^5 are performed. Data are collected at two different surface heat flux levels. The first is chosen to obtain a maximum surface temperature of approximately 100 °F, while the peak temperature at the higher heat flux is nearly 130 °F. Multiple numerator and denominator images are obtained at each heat flux level in order to assess the repeatability of the results. Two separate physical arrangements of the UV illumination source and the video camera are employed to collect data at radial positions from 37.5° to 120° from the leading edge stagnation point. The first equipment position spans the range 37.5° to 75°, while the second encompasses 82.5° to 120°.

RESULTS AND DISCUSSION

In the early phases of this study, the effects of variations in the UV illumination intensity, viewing angle and bandpass filter positioning were examined in a series of calibration tests. In these tests, a 1-inch (25 mm) square block of copper with a 1/4-inch (6 mm) diameter phosphor coating is employed. To permit precise surface temperature control, the rear surface of the block is immersed in an isothermal bath. A thermocouple embedded directly behind the phosphor dot is used to record the surface temperature. Test results indicate that the absolute value of the intensity ratio at a given temperature is dependent on both the illumination angle and intensity. Despite the variation of the absolute value of the intensity ratio, the functional relationship between the intensity ratio and temperature is found to be similar in the cases examined. Tests to assess the repeatability of optical surface temperature measurements reveal that the positioning of the bandpass filters is critical. With manual removal and replacement of the filters, intensity ratio values are observed to vary as much as 5 percent. With use of a filter holder in which the filters

are positioned by a slider mechanism, however, this variation in intensity ratio is virtually eliminated. As illustrated by the calibration curve in Figure 2 in which 4 sets of data are collected at each temperature, repeatability is excellent and accuracies of ± 1 °F appear feasible.

The variation of intensity ratio values as a result of changes in illumination or viewing angle is believed to be attributable to the energy absorption properties of the binder used to apply the thermographic phosphor. In earlier thermographic phosphor experiments with an optical heat flux gauge conducted at WL/POTC, a significant variation in the intensity ratio between phosphor coatings on the front and back surfaces of a thin, translucent insulator was observed. In these tests, the variation was determined to be the result of UV absorption by the insulator. Although the binder is suspected to have a similar effect, an explicit evaluation of the energy absorption properties of the binder material has not been conducted.

As is noted in the preliminary calibration tests, the intensity ratio at a given temperature is found to vary among the different radial positions on the cylinder as a result of differences in illumination angle and intensity. Despite these variations, individual calibration curves for each rectangular region exhibit an excellent fit with the calibration data, as illustrated in Figure 3. In general, the cylinder surface calibrations indicate that temperature measurements should be at least as accurate as ± 2 °F. Because the repeatability of the technique was established in preliminary calibrations, as discussed above, only one data point at each temperature is used in the cylinder surface calibrations.

While the results of preliminary and cylinder surface calibrations provide a measure of the accuracy and repeatability of this optical technique, the collection of experimental data provides a more robust validation of the technique. With the ability to control the heat flux from the foil surface of the cylinder, the measured circumferential temperatures were originally intended to be used to calculate local heat transfer coefficients that could be compared with previous empirical results. However, damage to the silver epoxy bond between the foil and brass disks resulted in an uneven surface heat flux, so it is not possible to accurately calculate local heat transfer coefficients. Instead, the measured circumferential temperatures are compared with the thermocouple data to assess measurement accuracy.

For the first positioning of the mercury vapor lamp and video camera, temperatures for radial positions from 37.5° to 75° from the leading edge stagnation point are measured.

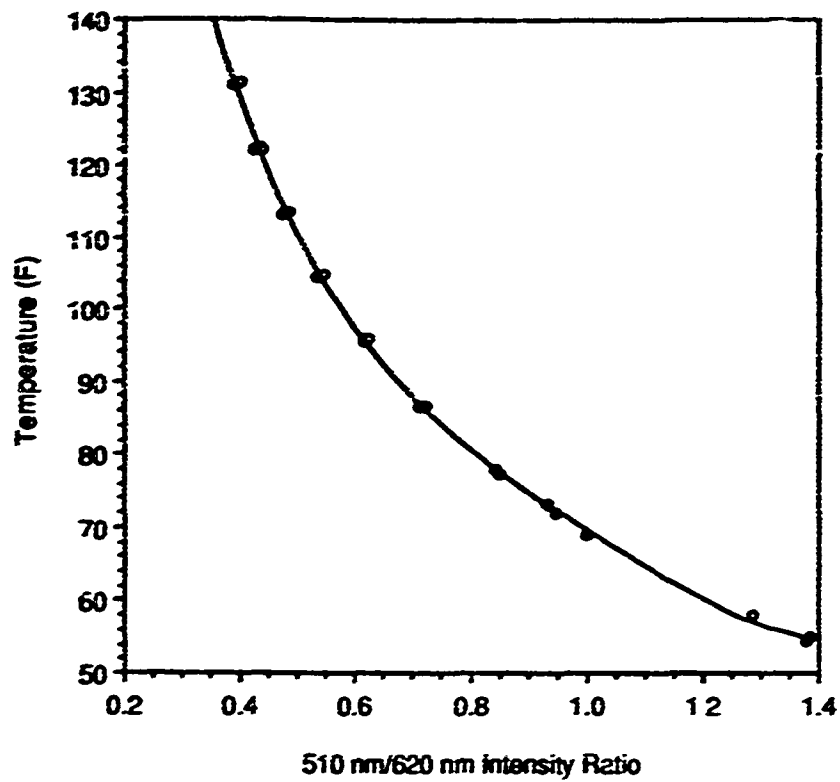


Figure 2 $\text{La}_2\text{O}_2\text{S}:\text{Eu}$ Phosphor Calibration

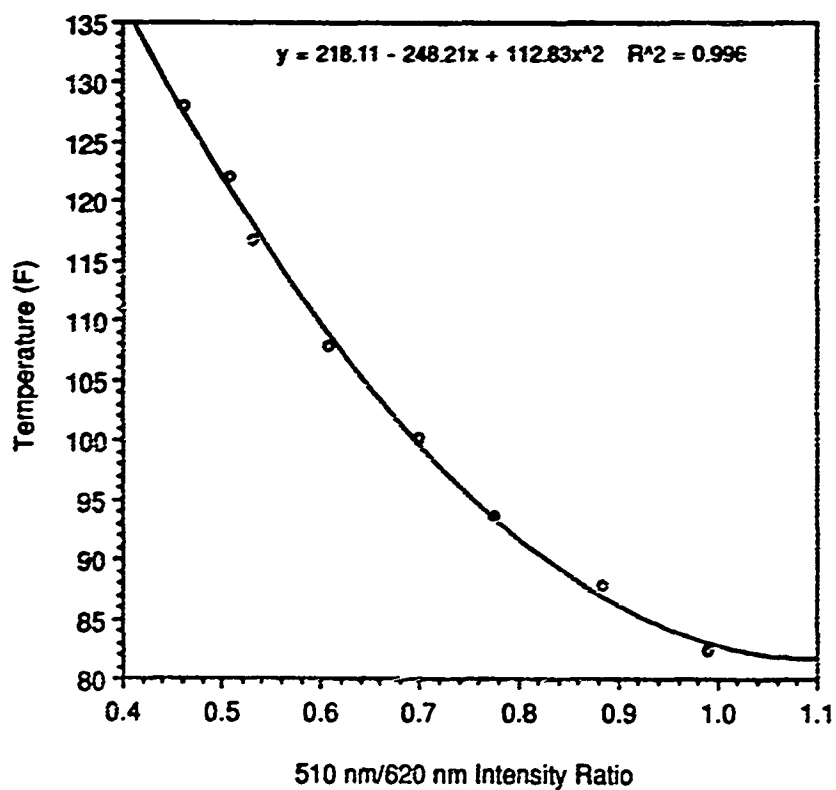


Figure 3 Calibration for Rectangular Region at 112.5° from Stagnation Point

Results for the low heat flux tests, which are presented in Figure 4, reveal measured temperatures that are lower than the thermocouple data by approximately 2 °F, or less. It is important to note, however, that the relatively large differences in thermocouple readings in these tests are the result of surface heating caused by the UV light source. In all other tests, surface heating was minimized by covering the lamp when data were not being collected. The results for the high surface heat flux tests, illustrated in Figure 5, track the thermocouple data quite well.

The second positioning of the mercury vapor lamp and video camera permits circumferential temperature measurement from 82.5° to 120° from the stagnation point. These data are the most interesting, since these radial positions permit identification of the separation point. Empirical values of the Nusselt number (Figure 6) indicate that the lowest temperature (i.e., the highest Nusselt number) should occur at approximately 115° from the stagnation point, while the highest temperature can be expected at a radial position slightly greater than 80° from the stagnation point. The results for the low- and high-heat flux tests agree well with these predictions.

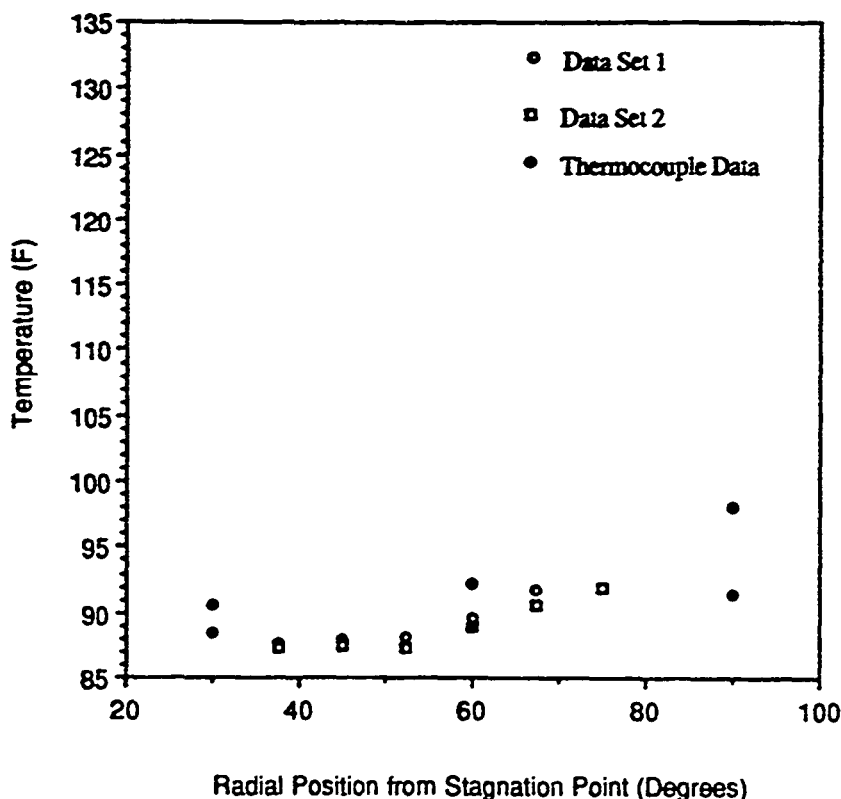


Figure 4 Low Heat Flux Test Results for Radial Positions from 37.5° to 75°

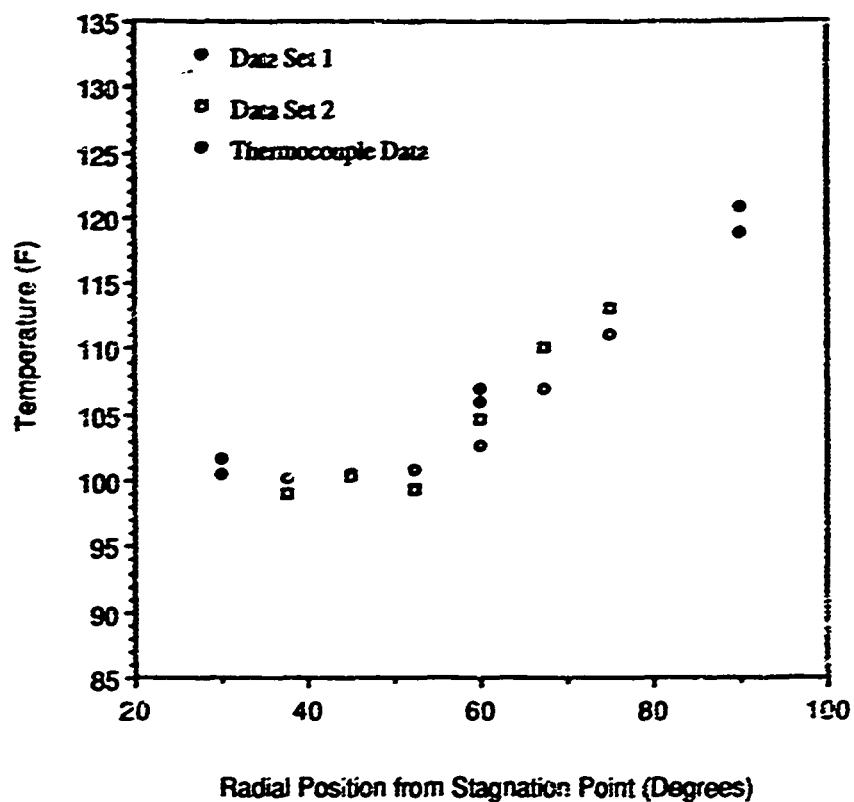


Figure 5 High Heat Flux Test Results for Radial Positions from 37.5° to 75°

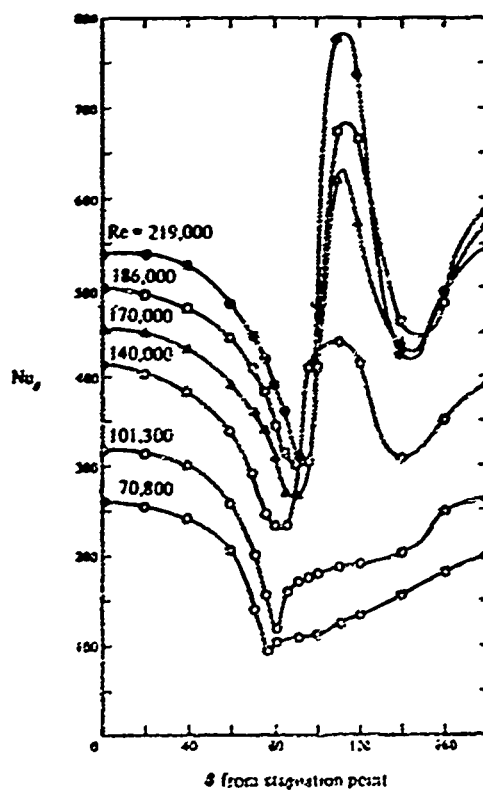


Figure 6 Local Nusselt Number Data for a Cylinder in Cross Flow

The low flux temperature data for the second lamp/camera position are presented in Figure 7. These results exhibit excellent repeatability and consistency with respect to the expected locations of maximum and minimum temperature. Also, the accuracy of the measured temperature when compared to the thermocouple data is $\pm 1.5^\circ\text{F}$, or better.

Results for the high heat flux tests, which are presented in Figure 8, are similar to those for the low surface heat flux. Surface temperature measurements are repeatable and the results are consistent with expectations. However, the measured surface temperature at 120° is approximately 6°F higher than the temperature obtained from the thermocouple mounted on the inner surface of the cylinder. Although unsettling, this disparity may be due to imprecise positioning of the thermocouple and/or lateral heat conduction. With the high circumferential temperature gradients indicated by the temperatures measured using the thermographic phosphor, a slight misplacement of the thermocouple in conjunction with lateral conduction in the epoxy bonding used to adhere the thermocouple to the foil could result in a false reading. The excellent repeatability of the optically-measured results and good agreement with the thermocouple data at the 90° position suggest that this is the case.

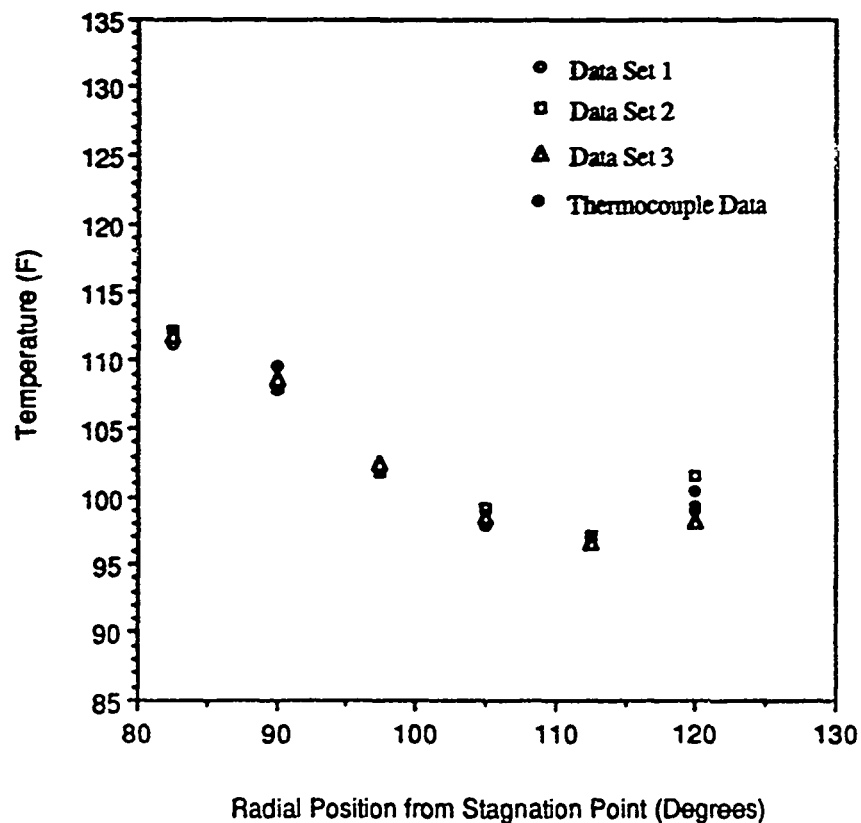


Figure 7 Low Heat Flux Test Results for Radial Positions from 82.5° to 120°

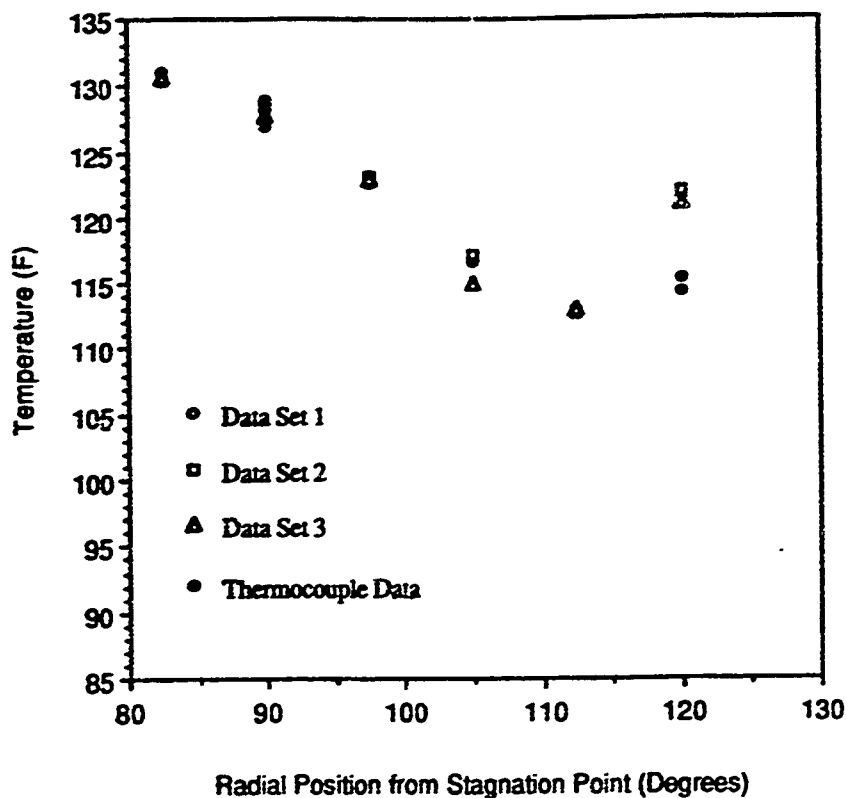


Figure 8 High Heat Flux Test Results for Radial Positions from 82.5° to 120°

CONCLUSIONS AND RECOMMENDATIONS

This experimental investigation of the circumferential temperature variation on a heated cylinder in cross flow has demonstrated the repeatability and accuracy of optical surface temperature measurement using thermographic phosphors. Despite the time constraints of this program, repeatable temperature measurements with an accuracy of ± 2 °F have been obtained. Furthermore, calibration tests have indicated that accuracies of ± 1 °F are feasible. Future system improvements, as discussed below, promise to further refine the technique so that it will be possible to obtain two-dimensional surface temperature measurements without requiring an in situ calibration of the phosphor surface.

A key limitation of the current setup is the limited output capability of the mercury vapor lamp. Although a more powerful vapor lamp would be expected to increase the surface area that can be effectively excited, use of a vapor lamp poses two difficulties. First, with continuous illumination of the test surface, localized surface heating may occur. Secondly, the output of a mercury vapor lamp is nonuniform, requiring diffusion to obtain a more uniform profile. While it is possible to address each of these concerns, an alternate solution is to excite the phosphor with a pulsed UV laser. The high power density (~75

mJ) and short duration pulse (5 ns) ensures adequate excitation with minimum surface heating. The drawback of this approach, however, is the need for an image-intensified CCD camera for image acquisition.

Another hardware component upgrade that might enhance the performance of the system is the image acquisition card. Currently, the analog intensity signal is digitized using 8 bits, which yields 256 intensity levels. In this experiment, the intensity level of the temperature sensitive emission line (510 nm) varies by approximately 12 intensity levels over a 50 °F range. The precision of intensity measurement could be significantly enhanced if a 10- or 12-bit digitization were used.

Finally, sputtering as a means of applying a phosphor film to a test surface should be examined. The sputtering process allows deposition of an extremely thin coating of pure phosphor. EG&G Measurements Inc. reports that the fluorescent intensities of sputtered surfaces are nearly as intense as those associated with the pure phosphor. More importantly, this application technique would eliminate the problem of UV and visible light absorption by a surface binder material.

Experimental and Analytical Investigation of Effects of Noncondensable Gases On On-Axis Rotating Heat Pipes

Kaveh A. Tagavi & Bryan T. Martin
Department of Mechanical Engineering
University of Kentucky
Lexington, Kentucky

Air Force Mentors:

Kirk L. Yerkes & Won S. Chang
Wright Laboratory
Dayton, Ohio

ABSTRACT

During the ten week summer appointment, we conducted an investigation of the effects of noncondensable gases on the performance of on-axis rotating heat pipes. Our efforts were divided in two fronts: experimental and analytical.

On the experimental front we designed and manufactured a rotating heat pipe made principally of oxygen-free-hard-copper with water as the working fluid. The assembly was isolated from the working table and electrical motor by using two air bearings. To study the effects of noncondensable gases some predetermined amount of nitrogen will be added to the working liquid. The heat pipe's inner wall was tapered at about two degree slope so that the centrifugal force would help the transfer of the working liquid from the condenser end to the evaporator end. Several thermocouples were installed both throughout the vapor passage and at the inner and outer sides of the heat pipe's wall. This will facilitate the measurement of temperatures of the heat pipe wall and the vapor inside. Accordingly, heat flux through the wall and the amount

of noncondensable gases may be calculated from the temperature measurements. The thermocouples are connected to a data acquisition system by a slip-ring. The condenser end, which is covered by radial fins, is cooled by the aid of a Vortex tube while the evaporator end is heated radiantly via a set of element heaters.

Although no systematic set of experiments was performed during this period, several sets of experiments are planned for the follow-up study during the next academic year.

On the analytical front, a dual approach was adopted. The vapor region, including the effect of noncondensable gases, was solved in an exact manner with a minimal number of simplifying assumptions. The liquid film, consisting of the working liquid in the condenser and evaporative regions, was analyzed in a manner similar to previous reports published by Marto et al. Then these two solutions were coupled through the interface condition. Of course, the coupling cannot be satisfied analytically in a way that would result in closed form solutions; rather, the coupling act has to be carried numerically. Again, the numerical solution of the analytical approach is planned for the upcoming academic year.

Finally, it is planned that the experimental and analytical results be compared with each other and with those available in the literature.

NOMENCLATURE

a	aspect ratio, L/r_1
C	total moles
c	molar concentration
D	diffusion coefficient
h	heat transfer coefficient in surrounding environment
h_{fg}	latent heat of vaporization
I_0	Bessel function of the first kind and zero order
I_1	Bessel function of the first kind and first order

K	property group, $(3k\Delta T\mu)/(h_{fg}\rho^2 r_1 \omega^2)$
k	liquid film thermal conductivity
L	axial length of heat pipe
N	molar flux
n	any integer
P_0	constant determining pressure field
p	pressure
R	dimensionless radial coordinate, r/r_1
R_c	universal gas constant
r	radial coordinate
r_1	inner radius of the heat pipe
T	temperature
T_i	interface temperature between liquid film and gas/vapor region
T_∞	temperature of surroundings
t	time coordinate
U	liquid velocity at condensation and evaporative interfaces
u	dimensionless liquid film velocity, Ur_1/D
V	molar velocity
x	molar fraction
y	$\ln(x_2)$
Z	dimensionless axial coordinate, z/L
z	axial coordinate

Greek Letters

β	variable, δ^4
δ	liquid film thickness
δ_0	liquid film thickness at $z=0$
λ	eigen value
μ	viscosity
ξ	variable, $d\beta/dz$

ρ	mass density
ϕ	angle of tapered inner wall
ω	angular velocity

Subscripts and Superscripts

1	vapor
2	noncondensable gas
c	condenser section
e	evaporator section
i	interface
∞	surrounding environment

INTRODUCTION

A rotating heat pipe (thermosyphon) utilizes the gravity field generated by rotation as a driving force for the movement of the working fluid from the condenser end to the evaporator end. This movement is due to the difference in hydrostatic pressure between the condenser and the evaporator (liquid film in the condenser side is generally thicker than that in the evaporator side). The difference in the hydrostatic pressure can further be enhanced by tapering the inside of the heat pipe.

Noncondensable gases exist in heat pipes either intentionally or unintentionally. Gases are generated in most heat pipes due to chemical processes such as corrosion. Noncondensable gases are also introduced in the heat pipe as a means of controlling the heat load ("Gas-Loaded Heat Pipes"). At any rate, even a small amount of noncondensable gases could affect the operation of the heat pipe considerably. As the working vapor moves from the evaporator end to the condenser end it also carries the noncondensable gases. Unlike the vapor, the noncondensable gases cannot condense and therefore will pile up at the condensation front. Thereafter, the pile-up of noncondensable gases will hinder the condensation process. For these reasons, it is

desirable to understand the behavior of the noncondensable gases, and how and where these gases build up (find the gas' distribution).

BACKGROUND

Since its invention in 1944, the heat pipe has been used in many situations that require transfer of heat with a minimum drop in temperature. A special kind of heat pipe (i.e. rotating heat pipe) is one that utilizes rotation to transport the working liquid from the condenser end to the evaporator end. An extensive survey of applications of rotating heat pipes is included in the paper by Yerkes (1990). Early on, it was realized that a greater thermal control of the heat pipe can be achieved by adding some predetermined amount of noncondensable gases. These gases will pile up in the condenser section, thereby reducing the effective area available for condensation. Gas-loaded heat pipes have been the subject of many studies. The simplest treatment of the noncondensable gases is reported by Dunn and Reay (1982). It simply assumes that there is a sharp wall separating the noncondensable gases from the vapor. This resulted in a quick calculation of the temperature profile along the heat pipe. Next, Edwards and Marcus (1972) treated the problem as one-dimensional and obtained the axial dependence of the noncondensable gas distribution. Later, Peterson and Tien (1989) used an integral technique to obtain an approximate, yet two-dimensional, profile for the noncondensable gas distribution. This was followed by Chang and Yu (1990) who showed how to get yet higher degree solutions.

Recently some attention has been devoted to rotating heat pipes. The idea is that the rotation can be used to transport the fluid from the condenser side to the evaporator side. This is achieved by tapering the inner wall. The liquid condensate flows for two reasons: hydrostatic pressure induced by rotation and the fact that the inner wall is tapered. Marto (1973) set out the foundation of treating the liquid flow in rotating heat pipes. He used the technique introduced by Leppert and Nimmo (1968). They dealt with laminar film condensation on horizontal surfaces where the only driving force is the hydrodynamic pressure differences.

Daniels and Al-Jumaily (1975) used a Nusselt-type approach for modeling the liquid film in a rotating heat pipe with no noncondensable gases. Later, Daniels and Williams (1978)

Baharnah (1980) studied rotating heat pipes with noncondensable gases. In their work they limited the effect to that of higher pressure due to the added gas and did not offer a rigorous analysis of the gas concentration profile.

EXPERIMENTAL EFFORTS

The most effective way to define the noncondensable gas front is to find the distribution of temperatures within the heat pipe. The reason for this is that as the working vapor becomes mixed with the noncondensable gases (in increasing concentrations toward the condenser end) the partial pressure of the working vapor will decrease, therefore changing the saturation temperature of the vapor. If we know how the temperature varies within the heat pipe, we will know how the partial pressure of the vapor changes. From this, we can then extrapolate the concentration of the noncondensable gas at various positions within the heat pipe. The experimental apparatus was designed with the use of this method in mind.

It is important, however, to keep in mind the effects of hydrostatic pressure, due to rotation, on the vapor pressure. In order to get some idea of the importance of this factor, a short analysis was done to approximate the increase in hydrostatic pressure at any given radius from the center of the heat pipe. The relation was found to be $p(\omega)/p(\omega=0) = \exp[(\omega^2 r_1^2)/(2R_c T)]$ using ideal gas assumptions, which were substantiated by the values this relation gives. Figure 1 shows the results of this relation for water vapor. As can be seen, this effect is not very large, even at very high rotational speeds and large radius. With this knowledge, we were satisfied that the saturation temperature of the vapor would be within reasonable limits. This gives us a means of correcting our data for the effects of rotation. Figure 2 is a plot of how this pressure difference will effect the temperature difference. It was composed with data from standard thermodynamic tables.

The test rig which we will use is needed by other researchers, so our apparatus had to be designed in modular form so it could be easily installed or removed when the time came for data collection. The test module is built around a foot-long thermosyphon cylinder. Figure 3a depicts the heat pipe while Figure 3b shows the overall test assembly. This cylinder has a two degree

internal taper to drive the working liquid, which is methanol. The condenser end has radial external fins to increase the wattage the condenser can handle. We decided against axial fins for several important reasons, which will not be detailed here. There are also several slots and holes machined into the exterior of the cylinder, these will accommodate thermocouple wires.

Each end has a series of flanges to facilitate easy dismantling for such things as refilling, cleaning, and servicing internal components. These are parts 2-7. Many are sealed with O-rings. Parts 2 and 4 have axial grooves, which are part of the thermocouple wire system. Part 6 is unique, with a diametral groove to channel the wires from the exterior of the module to the interior of the shaft, to be taken out of the module. Part 8 is the bearing interface, and is the piece which allows the apparatus to be truly modular.

Parts 9-11 connect together inside of the chamber to lend structural support to the "thermocouple tree" (part 11). Part 10 is the fill tube, and is welded to part 5. Part 9 is merely a coupler which allows 10 and 11 to connect with one another. The thermocouple tree is the means of getting temperature data from the inside of the thermosyphon. It is a straight copper tube with two rows of branches along its length at 180 degrees. This serves as a conduit for thermocouple wires inside the thermosyphon. Various configurations of branches will be used in order to accomplish the goal of mapping the temperature distribution within the heat pipe.

Temperature data is taken out of the rotating module by means of a slip ring assembly. The wires are channelled into the shaft (part 14), through the bearing, through the slip ring adaptor (13) and finally to the slip ring itself. The readings are then recorded on a Fluke data acquisition unit.

Heat is supplied to the system by two clam-shell heaters, which will be contained in an insulated heater housing. They have a maximum heating capacity of 800 W, which is high enough for methanol.

Cooling at the condenser end will be accomplished with cold air. A Vortex tube forces cold air into a small chamber, which is within a plexiglass cooling box (similar in design to the heater housing).

There are three ways to estimate the heat load of the system. The first is simply to note the amount of power supplied to the heater. The second is to measure the total heat flux at the evaporator wall. The thermocouples are embedded at different depths in the evaporator wall in order to establish a temperature gradient. This will give an estimate of the conductive heat flux. Lastly, and probably the most accurate, is the measure of heat convected out by the air flow in the condenser. Flow-rate will be measured and air temperatures monitored to get an overall rate of heat removal. We are in the process of calibrating the slip ring for inaccuracies due to temperature and rotation speeds of the slip ring itself.

The test will be driven by a 2 hp electric motor through a tuned shaft. We are therefore limited, at the moment, to conducting tests at fairly low speeds (up to 3000 rpm). It is our hope to eventually acquire the means to run our test at much higher speeds (20,000) to better simulate aircraft operating conditions. It was for this reason that our apparatus was made much stronger than is necessary for our tests. Data will be taken while varying several different parameters; these include rotational speed, amount of noncondensable gases, and heat load.

Due to a great deal of out-of-house machining, and the failure to get the air bearings shipped in time, the test is not yet assembled. This will be accomplished over the course of the coming academic year. Bryan Martin will work on this project as the topic of his masters thesis and will maintain close contact with this laboratory, making return trips whenever possible.

ANALYTICAL EFFORTS

As mentioned before, the analytical approach consists of modeling the vapor/gas region and the liquid film separately. Later the two solutions are coupled through the interfacial matching condition. Figure 4 shows the two different regions.

VAPOR/GAS SOLUTION

In the absence of chemical reactions, the equation governing the molar concentration is written as

$$\frac{\partial c_2}{\partial t} + \nabla(N_2) = 0 \quad (1)$$

where c is molar concentration, t is time, N is the molar flux, and subscripts 1 and 2 represent vapor and gas. Assuming steady state, the governing equation reduces to

$$\nabla(N_2) = 0 \quad (2)$$

Since N_2 is zero at all boundaries, it can be shown that the only acceptable solution is

$$N_2 = 0 \quad \text{everywhere} \quad (3)$$

The molar flux N_2 , however, could be related to the mixture velocity as

$$N_2 = x_2 c V - c D \nabla(x_2) \quad (4)$$

where x is molar fraction, V is velocity, D is diffusion coefficient, and variables with no subscript represent those for the vapor/gas mixture. Equations (3) and (4) may be combined to solve for V as

$$V = D \nabla(\ln x_2) \quad (5)$$

This equation and the continuity equation

$$\nabla(V) = 0 \quad (6)$$

result in a conduction-type equation that governs the molar fraction for the gas.

$$\nabla^2(y) = 0 \quad (7)$$

where a new variable has been defined, mainly for convenience, as

$$y = \ln x_2 \quad (8)$$

Finally, the molar fractions for the gas and vapor are related as

$$x_1 + x_2 = 1 \quad (9)$$

Equation (8) may be solved to obtain molar fractions for the vapor and gas. Having the molar concentrations, then equation (9) may be used to obtain the molar velocity of the mixture at any point. Next, assuming inviscid flow, the pressure of the mixture is obtained from potential flow considerations as

$$p = P_0 - \frac{\rho_1 v^2}{2} \quad (10)$$

where P_0 is an arbitrary number. The determination of the constant P_0 will be discussed later.

BOUNDARY CONDITIONS

Figure 5 shows a schematic of the vapor/gas region including the coordinates. The vapor molar flux may be written as

$$N_1 = x_1 cV - cD\nabla x_1 \quad (11)$$

The molar velocity is related to the molar fluxes as

$$cV = N_1 + N_2 \quad (12)$$

Substitution of equation (12) into equation (11), considering that $N_2=0$, and rearranging results in

$$N_1 = -\frac{cD\nabla x_1}{1 - x_1} \quad (13)$$

Or in terms of x_2

$$N_1 = \frac{cD\nabla x_2}{x_2} = cD\nabla y \quad (14)$$

At the evaporative boundary, we can write

$$N_1 = cD\nabla y = -cU_e \quad (15)$$

or

$$\nabla y = \left(\frac{\partial y}{\partial r}\right)_{r=r_1} = -\frac{U_e}{D} \quad (16)$$

Similarly at the condensation boundary,

$$\left(\frac{\partial y}{\partial r}\right)_{r=r_1} = \frac{U_c}{D} \quad (17)$$

At all other boundaries the condition is that of no mass transfer, that is

$$\left(\frac{\partial y}{\partial r}\right)_{r=r_1} = 0 \quad (18)$$

The symmetrical boundary condition at $r=0$ is written as

$$\left(\frac{\partial y}{\partial r}\right)_{r=0} = 0 \quad (19)$$

The velocities U_c and U_e are related to the rate of cooling and heating at the condenser and evaporative sections. Figure 6 shows the schematic of a simple model that describes these sections. The rate of heat transfer can be related to the velocities U as

$$\rho_c U_c h_{fg} = \frac{T_{i,c} - T_{o,c}}{\frac{1}{h_c} + \frac{\delta_c}{k_c}} \quad (20)$$

$$\rho_e U_e h_{fg} = \frac{T_{o,e} - T_{i,e}}{\frac{1}{h_e} + \frac{\delta_e}{k_e}} \quad (21)$$

where h_{fg} is the latent heat of vaporization, T_i is the temperature of the liquid film interface, T_o is the surrounding temperature, h is the outside heat transfer coefficient, δ is the liquid film thickness, and k is the liquid thermal conductivity.

SOLUTION FOR y

The governing equation (7) and boundary conditions (16-19) can, in principle, be solved by separation technique. Assuming circumferential symmetry, the governing equation reduces to.

$$\frac{\partial^2 y}{\partial z^2} + \frac{1}{r} \left(r \frac{\partial y}{\partial r} \right) = 0 \quad (22)$$

We may switch to a dimensionless set of variables such that

$$Z = z/L \quad R = r/r_1 \quad a = L/r_1 \quad (23)$$

The dimensionless governing equation can be written as:

$$\frac{\partial^2 y}{\partial Z^2} + a^2 \frac{1}{R} \frac{\partial}{\partial R} \left(R \frac{\partial y}{\partial R} \right) = 0 \quad (24)$$

The variable y can be separated such that

$$y(R, Z) = y_R(R) y_Z(Z) \quad (25)$$

The governing equation may now be separated as

$$\frac{1}{y_Z} \frac{d^2 y_Z}{dZ^2} = - \frac{a^2}{y_R} \frac{1}{R} \frac{d}{dR} \left(R \frac{dy_R}{dR} \right) = -\lambda^2 \quad (26)$$

with the following homogeneous boundary conditions

$$\frac{\partial y_Z}{\partial Z} = 0 \quad \text{at } Z=0 \text{ and } 1 \quad (27)$$

and

$$\frac{dy_R}{dR} = 0 \quad \text{at } R=0 \quad (28)$$

The separate solutions are

$$y_Z = (\text{const.}) \cos(\lambda Z) \quad (29)$$

where the eigenvalue λ is obtained from

$$\lambda = n\pi \quad n = \text{any integer} \quad (30)$$

and

$$y_R = (\text{const.}) I_0\left(\frac{\lambda}{a} R\right) \quad (31)$$

Therefore

$$y = \sum_{n=1}^{\infty} A_n I_0\left(\frac{n\pi}{a} R\right) \cos(n\pi Z) \quad (32)$$

The constant A is calculated by satisfying the interface condition at the condenser and evaporative sections. The interface condition is written as

$$\left(\frac{\partial y_R}{\partial R} \right)_{R=1} = u(Z) \quad (33)$$

where

$$u(Z) = -\frac{U_{c1}}{D} \quad \text{for } 0 < Z < \frac{L_e}{L} \quad (34)$$

$$u(Z) = 0 \quad \text{for } \frac{L_e}{L} < Z < 1 - \frac{L_c}{L} \quad (35)$$

$$u(Z) = \frac{U_{c1}}{D} \quad \text{for } 1 - \frac{L_c}{L} < Z < 1 \quad (36)$$

The remaining task is to expand u in terms of a series of Bessel functions.

$$u(Z) = \sum_{n=1}^{\infty} 2 \left\{ \int_0^1 u(\zeta) \cos(n\pi\zeta) d\zeta \right\} \cos(n\pi Z) \quad (37)$$

The constant A can be found by comparing equations (37) and (32) and substituting in equation (32) to get the final result for y as

$$y(R, Z) = - \sum_{n=1}^{\infty} \frac{2 \int_0^1 u(\eta) \cos(n\pi\eta) d\eta}{\frac{n\pi}{a} I_1\left(\frac{n\pi}{a}\right)} I_0\left(\frac{n\pi}{a} R\right) \cos(n\pi Z) \quad (38)$$

Since all the boundary conditions resulting in equation (38) were of the second kind, an arbitrary constant could be added to the solution. Using equation (8), the molar ratio for the gas may be written as

$$x_2(R, Z) = (\text{const.}) e^{y(R, Z)} \quad (39)$$

This constant may be obtained by checking the inventory of the noncondensable gases

$$2\pi r_1^2 L \int_0^1 \int_0^1 (\text{const.}) e^{y(R, Z)} c dR dZ = C_2 \quad (40)$$

where C_2 is the total number of moles of the noncondensable gas in the heat pipe. Calculating the constant from equation (39), replacing c by $p/R_c T$ where R_c is the universal gas constant, and substituting it in equation (38) yields

$$x_2(R, Z) = \frac{C_2 e^{y(R, Z)}}{2\pi r_1^2 L \int_0^1 \int_0^1 e^{y(R, Z)} \left(\frac{p}{R_c T}\right) dR dZ} \quad (41)$$

where y is calculated from equation (38).

SOLUTION FOR LIQUID FILM

The governing equation relating the rate of heat transfer to the rate of change in the liquid film thickness is taken from Marto et al., and is written as

$$\frac{k(r_1 + z \sin \phi) \Delta T}{\delta h_{fg}} = \frac{d}{dz} \left\{ \frac{\rho^2 \omega^2}{\mu} (r_1 + z \sin \phi) \left(\sin \phi - \cos \phi \frac{d\delta}{dz} \right) \frac{\delta^3}{3} (r_1 + z \sin \phi) \right\} \quad (42)$$

where variables without subscripts are now those of the liquid, ϕ is the angle of the tapered inner wall, ω is the angular velocity of the heat pipe, and μ is the viscosity. Assuming $r_1 \gg z \sin \phi$, equation (42) may be rearranged to

$$\delta \frac{d}{dz} \left\{ \left(\sin \phi - \cos \phi \frac{d\delta}{dz} \right) \delta^3 \right\} = K \quad (43)$$

where K is a dimensionless group defined as

$$K = \frac{3k\Delta T\mu}{h_{fg}\rho^2 r_1 \omega^2} \quad (44)$$

This equation (42) may be rearranged as

$$\frac{3}{4} \sin \phi \frac{d\beta}{dz} - \frac{1}{4} \cos \phi \beta^{1/4} \frac{d^2\beta}{dz^2} = K \quad (45)$$

where $\beta = \delta^4$. This equation cannot be solved in its present form. However, a change of variable such as

$$\xi = \frac{d\beta}{dz} \quad (46)$$

will make a solution viable. The new form of equation (45) is written as

$$\sin \phi \xi - \frac{1}{8} \cos \phi \frac{d\xi^2}{d\beta^{3/4}} = K \quad (47)$$

Rearranging equation (47) results in

$$d(\beta^{3/4}) = -\frac{\cos\phi}{4} \frac{\xi d\xi}{\frac{4}{3}K - \sin\phi \xi} \quad (48)$$

This equation can be integrated to solve for the liquid film thickness. The boundary conditions for δ are

$$\frac{d\delta}{dz} = \tan\phi, \quad \delta = \delta_0 \quad \text{at} \quad z=0 \quad (49)$$

The first condition is derived from the fact that at $z=0$ there would be no liquid flow. These two conditions in terms of ξ and β become a single equation, which is

$$\beta = \delta_0^4 \quad \text{at} \quad \xi=0 \quad (50)$$

Equation (48) now can be integrated using boundary condition

$$\delta^3 = \delta_0^3 + \frac{\cos\phi}{4} \left\{ \frac{\xi}{\sin\phi} + \frac{4K}{3\sin^2\phi} \ln\left[\frac{4/3K - \sin\phi \xi}{4/3K}\right] \right\} \quad (51)$$

A closed form solution in terms of δ and z cannot be obtained. Once δ vs ξ is obtained from equation (51), ξ may be related to δ and z from

$$\xi = 4\delta^3 \frac{d\delta}{dz} \quad (52)$$

This is a versatile equation: it can be used with heating (evaporative free surface), cooling (condensing free surface), or insulated wall (no mass transfer at the free surface), by using a positive, negative, or 0 value for ΔT .

SOLUTION METHODOLOGY

In the previous sections, analytical solutions for the vapor/gas region and liquid film have been obtained. These two solutions are related through the thermodynamics of the interface. An overall closed form and analytical solution is not possible due to the complex nature of the analytical results and the thermodynamic relation that relates the local molar fraction of the vapor to the vapor/liquid film interface temperature. Numerical results, however, can easily be obtained.

An iterative procedure is unavoidable, due to the nature of the nonlinearities in the problem.

One possible iterative scheme is listed below:

- 1) Guess an initial U_c and U_e , realizing that to achieve steady state the mass balance dictates that the vaporization mass should be equal to the condensing mass.

$$\int_0^{L_e} \rho_e U_e dz = \int_{L-L_c}^L \rho_c U_c dz \quad (53)$$

- 2) Calculate $u(Z)$ from equations (33)-(35).
- 3) Calculate $y(R,Z)$ from equation (38).
- 4) Calculate x_2 from equation (41).
- 5) Calculate x_1 from equation (9).
- 6) Having x_2 , the partial pressure of the vapor may be calculated from $p_2 = x_2 p$.
- 7) The liquid film-vapor interface temperatures $T_{i,c}$ and $T_{i,e}$ may now be calculated from thermodynamic tables.
- 8) The temperature difference may then be calculated as

$$\Delta T = T_{\infty,e} - T_{i,e} \quad \text{or} \quad \Delta T = T_{i,c} - T_{\infty,c} \quad (54)$$

- 9) The liquid film thickness δ is calculated from equation (51).
- 10) The velocities U_c and U_e can now be calculated from equations (20) and (21).
- 11) Check if calculated values of U_c and U_e are close to the guessed values? If yes, stop the iterations; if not, go back to step 1.

FUTURE DIRECTIONS

This subject has barely been touched. The combination of noncondensable gases and rotation has not been fully investigated before. The same combination is also responsible for the complex and interesting phenomenon at hand. Since no extensive result has been obtained here due to limited time and the scope of the project, more work will need to be planned for the future. Among the activities proposed for the future are:

- 1) Perform analysis for the case of no tapering. Because the heat pipe might be rotating very fast (up to 30,000 rpm) it is suspected that tapering is unnecessary for liquid transport. A thicker liquid film at the condensate section and a high centrifugal force might generate enough hydrostatic head to supply the necessary liquid flow.
- 2) Following the analysis in (1) above, numerical results should be obtained and presented in terms of plots and tables.
- 3) Numerical results will be obtained for the concentration of the noncondensable gas. These results will be coupled with those obtained by Edwards, Tien.
- 4) Overall results will be obtained in which the solutions of the vapor/gas region and the liquid film are coupled. These results will be compared with those available in the literature -- possibly those of Al-Jumaily (1973) and Williams (1976).
- 5) Experiments will be conducted. The data will be compared with those results obtained here and those by Edwards and Marcus (1972) and Peterson and Tien (1989).
- 6) A comprehensive computer program which accepts a few input parameters such as angular revolution, system pressure, and amount of noncondensable gases and solves for rate of heat transferred and the temperature profile in the heat pipe is desirable.

- 7) Upon examination of the experimental data and analytical results, recommendations will be made for better designs and more efficient analytical treatment.
- 8) The possible effect of ultra high revolutions on the performance of the gas-loaded heat pipes is not that clear at this point. Future experimental efforts aimed at investigating this are recommended.

REFERENCES

Al-Jumaily, F. K., "Investigation of the Factors Affecting Performance of a Rotating Heat Pipe," Ph.D. Thesis, University College of Swansea, Wales, 1973.

Chang, W. S., and Yu, J. S., "A note on the Gas Distribution in a Cylindrical Gas-Loaded Heat Pipe," *Journal of Heat Transfer*, Volume 12, 1990, pp. 779-781.

Daniels, T. C., and Al-Baharnah, N. S., "Temperature and Heat Load Distribution in Rotating Heat Pipes," *ALAA Journal*, Volume 18, Number 2, 1980, pp. 202-207.

Daniels, T. C., and Al-Jumaily, F. K., "Investigations of the Factors Affecting the Performance of a Rotating Heat Pipe," *International Journal of Heat and Mass Transfer*, Volume 18, 1975, pp. 961-973.

Daniels, T. C., and Williams, R. J., "Experimental Temperature Distribution and Heat Load Characteristics of Rotating Heat Pipes," *International Journal of Heat and Mass Transfer*, Volume 21, 1978, pp. 193-201.

Daniels, T. C., and Williams, R. J., "The Effect of External Boundary Conditions on Condensation Heat Transfer in Rotating Heat Pipes," *International Journal of Heat and Mass Transfer*, Volume 22, 1979, pp. 1237-1241.

Edwards, D. K., and Marcus, B. D., "Heat and Mass Transfer in the Vicinity of the Vapor-Gas Front in a Gas-Loaded Heat Pipe," *Journal of Heat Transfer*, Volume 94, 1972, pp. 155-162.

Leppert G, and Nimmo, B., "Laminar Film Condensation on Surfaces Normal to Body or Inertial Forces," *Journal of Heat Transfer*, Volume xx, 1968, pp. 178-179.

Marto, P. J., "Laminar Film Condensation on the Inside of Slender, Rotating Truncated Cones," *Journal of Heat Transfer*, Volume xx, 1973, pp. 270-272.

Petreson, P. F., and Tien, C. L., "Numerical and Analytical Solutions for Two-Dimensional Gas Distribution in Gas-Loaded Heat Pipes," *Journal of Heat Transfer*, Volume 111, 1989, pp. 598-604.

Williams, R. J., "Influence of Inert Gases on the Performance of Rotating Heat Pipes," Ph.D. Thesis, University of College Swansea, Wales, (1976).

Yerkes, K. L., "Technology Review: Utilizing Rotating Thermosyphon Technology in Aircraft Thermal Management and Control," *Aerospace Technology and Exposition*, Long Beach, California, October 1-4, 1990.

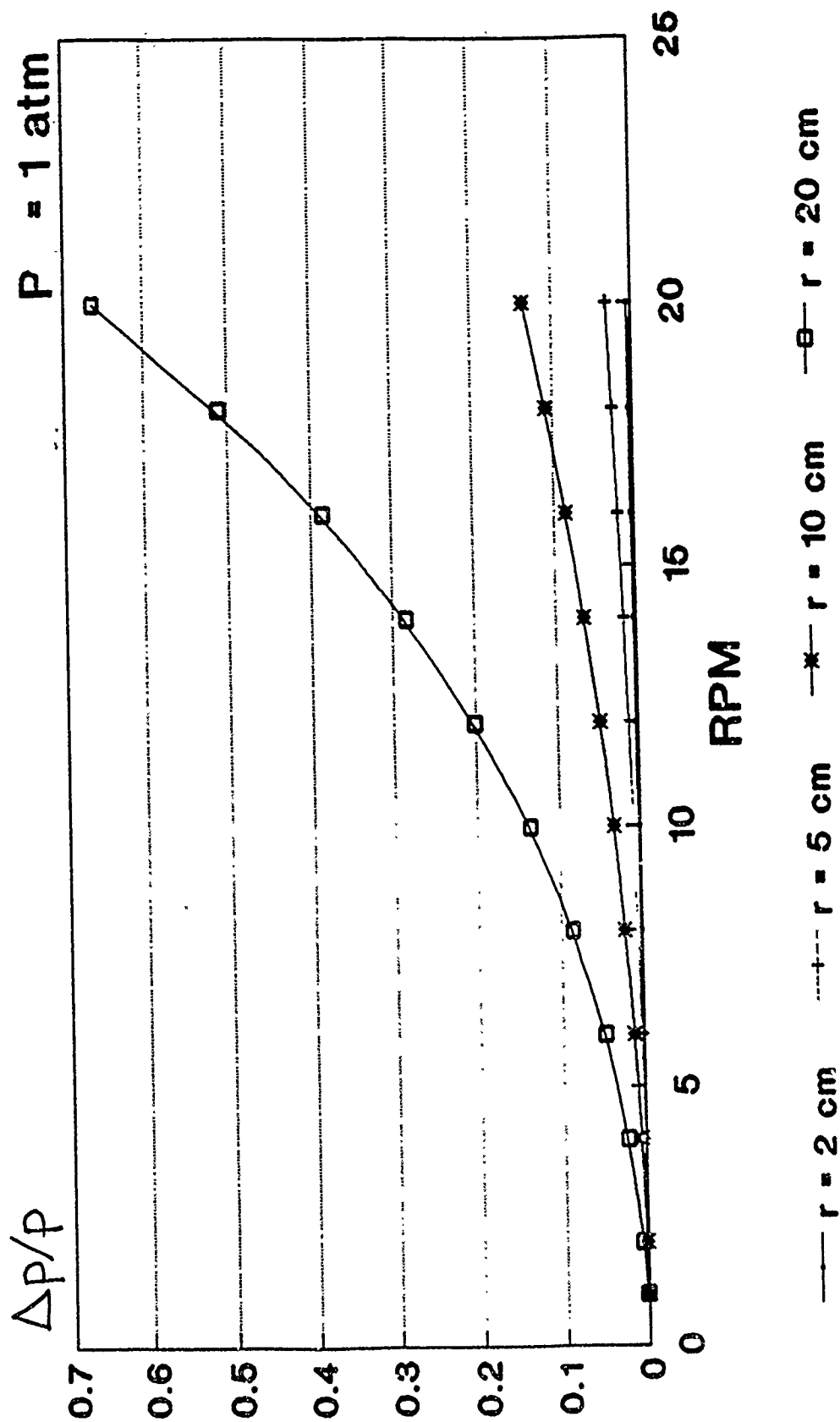


Figure 1.

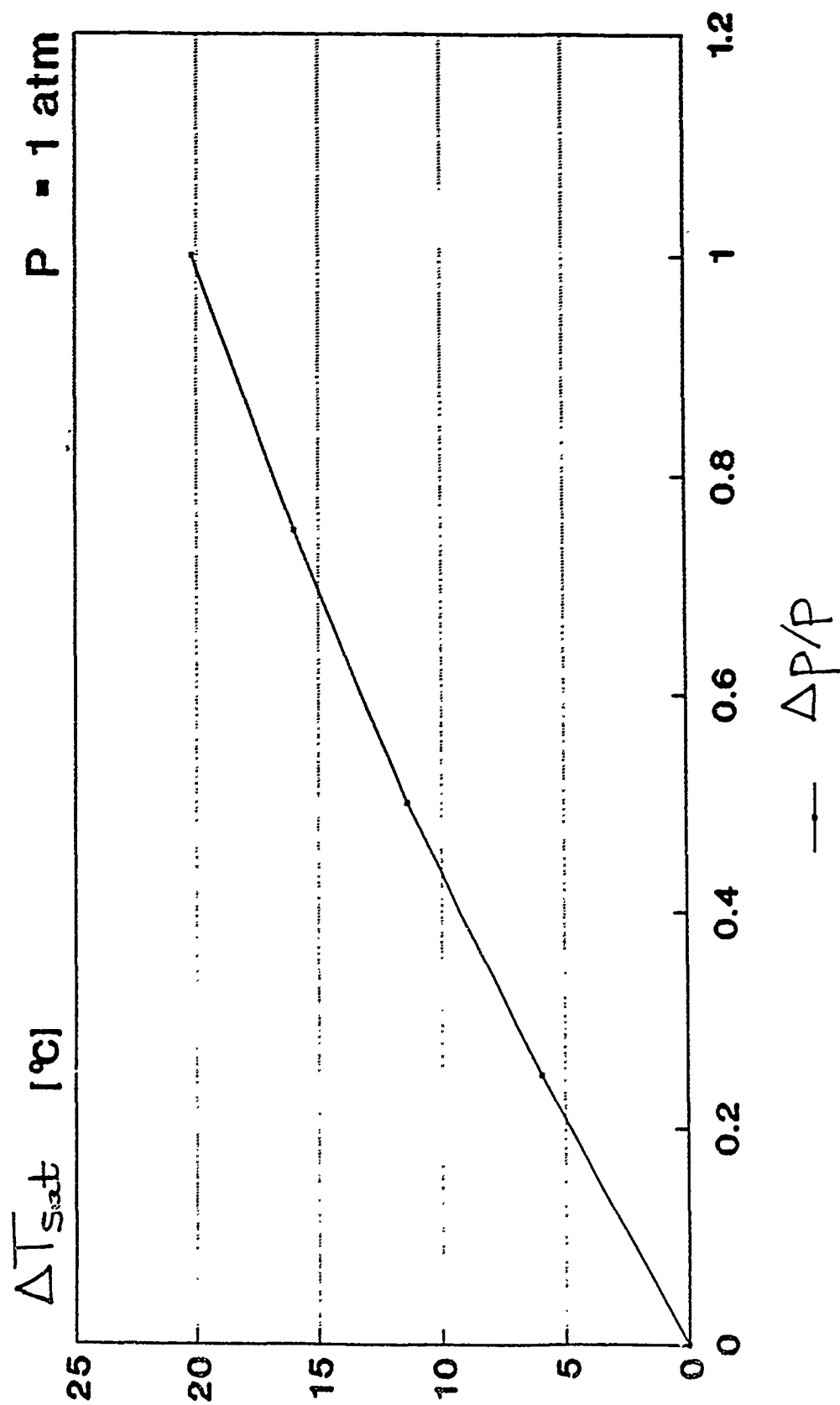
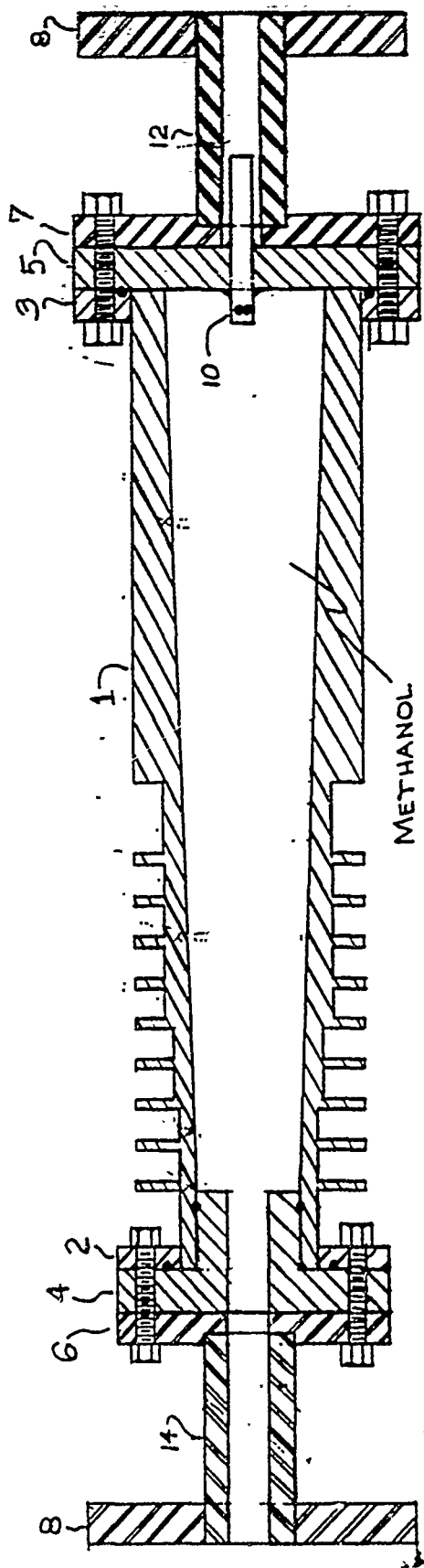
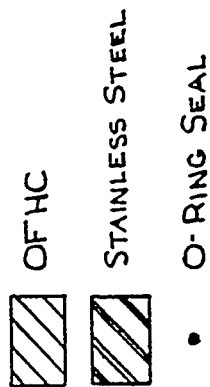
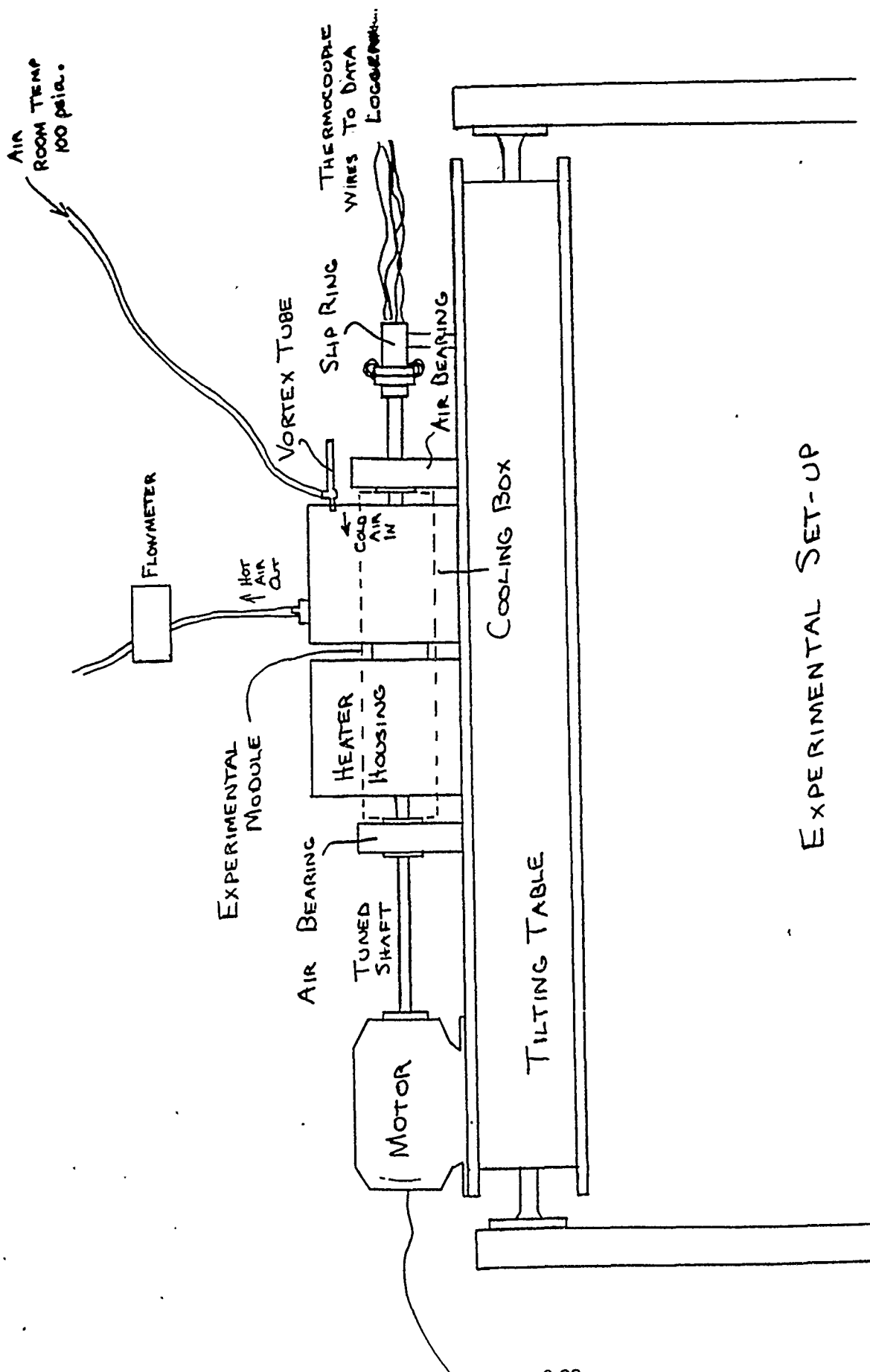


Figure 2.



TEST MODULE...
ROTATING THERMOSYPHON

Figure 3a



EXPERIMENTAL SET-UP

Figure 3b

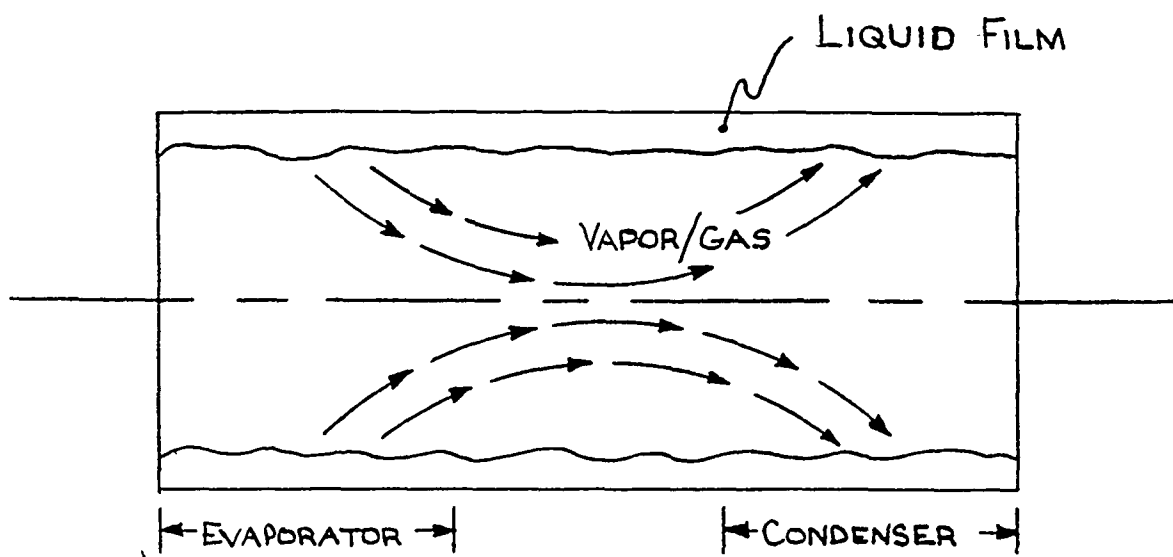


FIGURE 4

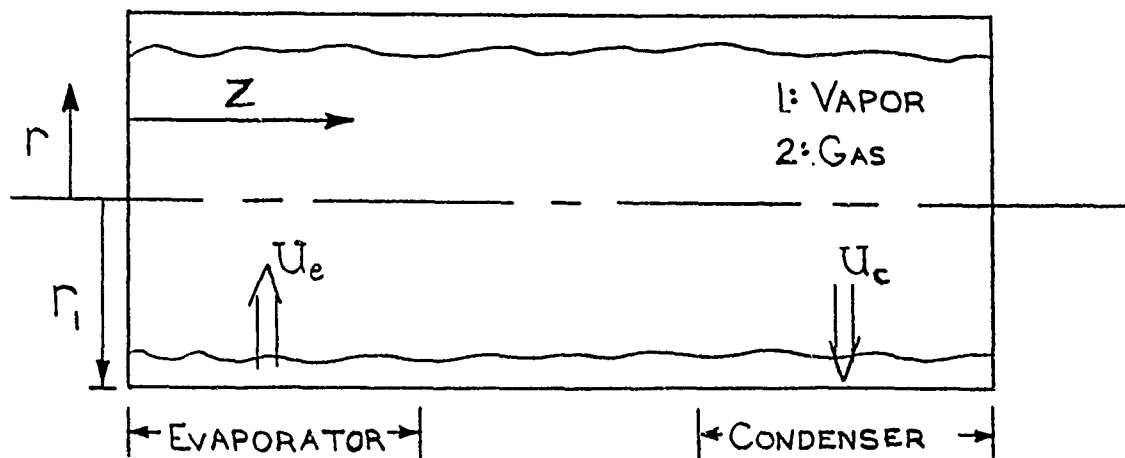


FIGURE 5

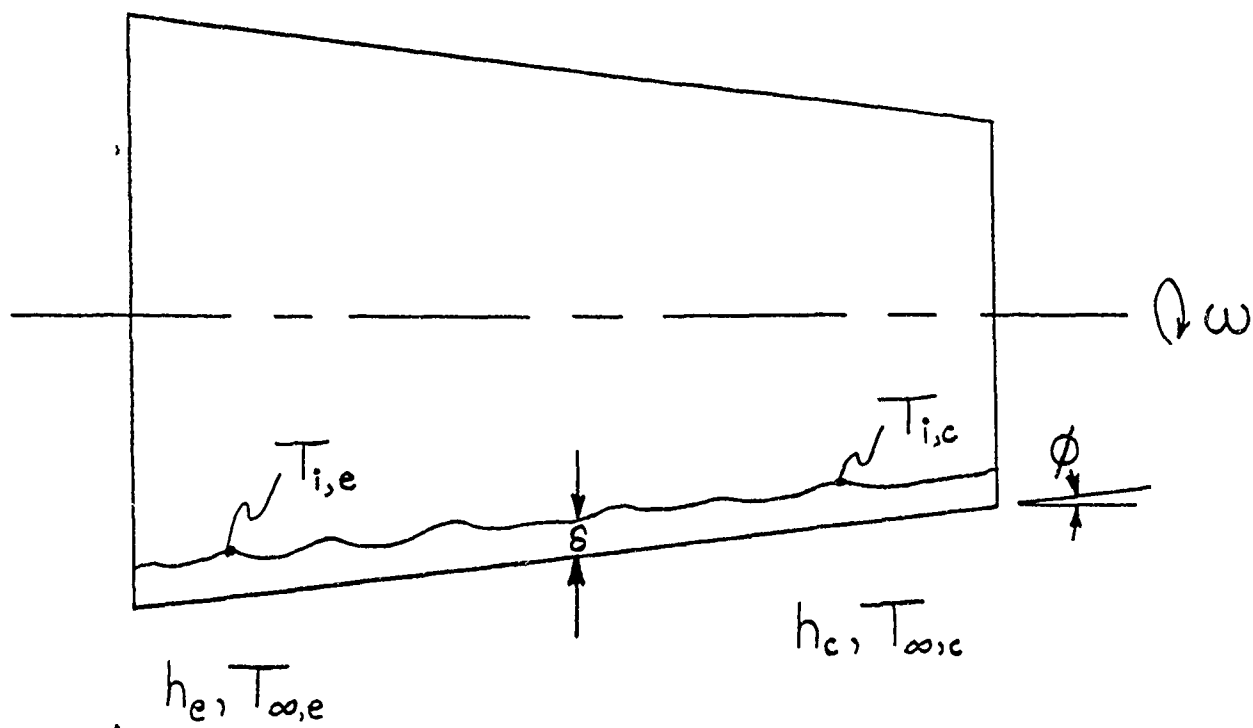


FIGURE 6

A TWO-WIRE THERMOCOUPLE TECHNIQUE FOR DYNAMIC TEMPERATURE MEASUREMENT IN REACTING FLOWS

by

Thomas R. Scattergood

ABSTRACT

A new method of determining the instantaneous thermocouple time constant is examined using two thermocouple junctions of similar composition but differing diameters placed in essentially the same environment. This method is found to work well in theory but is somewhat more complicated in practice due to a strong sensitivity of the technique on the ratio of junction sizes and on the nature of the response signal. A probe calibration technique is recommended and discussed for use with this method. Also, the effect of probe geometry on the flowfield was examined using an LDV in a simple dump combustor and it was determined that there were no significant effects caused by any of the three geometries studied.

INTRODUCTION

When studying combustion systems and reacting flows, it is important to be able to measure spatially and temporally resolved temperatures. This data, along with corresponding velocity data, aides in the understanding of various aspects of fluid flow such as stability characterization and control, interpretation of velocity data, and the understanding of turbulence correlations and compressibility effects. Since these fluid phenomena are of a rapidly changing nature, it is important to have a temperature measurement rate sufficient to keep up with such fluctuations in the flowfield. In this case, a thermal data rate of about 1 kHz would be required for the anticipated application to research ramjet combustors.

Current optical techniques based on the quantum-mechanical effects of light are useful in their ability to provide temperature measurements in flows due to their non-intrusiveness, their accuracy, and their speed. They can, however, be prohibitively expensive and very complicated, and none can yet provide combined spatial and temporal resolution at a satisfactory sampling rate. Thermocouples, on the other hand, are relatively cheap and uncomplicated and have a great deal of past experience behind them on which to rely upon. There still remain problems involving probe survivability and disturbance, as well as the usual worries over data validity, but the current research is attempting to provide insight into these areas of concern.

Thermocouples are able to provide very accurate information about the temperature at the probe junction, but there is a relationship between the junction temperature and the instantaneous gas temperature. Since the physical size of any thermocouple prevents it from responding directly to temperature fluctuations at frequencies above about 200 Hz, compensation is necessary. For the case of minimal radiation and conduction losses, it may be shown that:

$$T_c = T_j - \alpha (dT_j/dt)$$

where T_c is the actual gas temperature, T_j is the measured junction temperature, and α is referred to as the time constant, which represents a ratio between the junction thermal capacity and the convective heat flux. Since α is a function of the instantaneous gas properties, calculating the time

constant directly would not normally be possible, and most researchers have used some sort of average value. This has proven to be insufficient and an alternate method of determining α , and thus the instantaneous gas temperature, is needed.

The solution proposed here is to use a two-wire thermocouple in which two junctions of differing diameters are placed close together. This technique, theoretically, allows the determination of an instantaneous time constant based on the temperature gradients of each junction. The focus of this work is to prove this technique and apply it to an actual flow with possible future correlation to a CARS system and then to an LDV for simultaneous velocity measurements.

THEORY

Performing an energy balance on a thermocouple junction and neglecting radiation and conduction effects gives the standard equation for response:

$$T_c = T_j - \alpha (dT_j/dt)$$

where T_c is the instantaneous gas temperature, T_j is the junction temperature, and the time constant $\alpha = (\rho_j C_j d_j^2) / (4 k Nu)$. For small wires in turbulent flows, $Nu = C_1 Re^{C_2} Pr^{C_3}$. Substituting relations for Nu , Re , and Pr into the relation for α :

$$\alpha = [(\rho_j C_j d_j^2 - C_2) / 4] \times [\mu / (V^{C_2} \rho^{C_2} Pr^{C_3} C_1 k)]$$

Which may be grouped as

$$\alpha = (C_4)(G)$$

where $C_4 = C_4(\text{wire properties})$ and $G = G(\text{gas properties})$. Now,

$$T_c = T_j + C_4 G (dT_j/dt)$$

Clearly, G is the only unknown in solving for T_c . However, for two thermocouples in same environment, we can let $T_{c1} = T_{c2}$ and solve for G in terms of known or measurable parameters:

$$G = [T_{j2} - T_{j1}] / [(C_4 \, dT_j/dt)_1 - (C_4 \, dT_j/dt)_2]$$

With a known value for G , the time constant can now be solve for and an instantaneous gas temperature obtained.

This is a dynamic calibration and depends on the responses of two different junction in the same environment. The only required information is the value of the coefficient C_2 , which has been shown to be about 0.5 but can be experimentally verified.

In order to show that it is possible to obtain a compensated temperatures based on the technique just outlined, a theoretical response curve was generated to test this technique. By letting $T_c = T_0 \sin \omega t$ and substituting back into the relation $T_c = T_j + \alpha (dT_j/dt)$ an expression for the response of each junction T_j could be obtained:

$$T_j(t) = T_0 \alpha \times [(\sin(\omega t) / \alpha) - (\omega \cos(\omega t)) - \omega e^{(-t/\alpha)}] / [1 + (\omega t)^2]$$

By assuming a junction diameter ratio of 1:2 and assigning arbitrary values to T_0 , ω , and the other physical variables contained within α , the response curve for each junction was calculated. This can be seen in Fig. (1) where the curve of largest amplitude is some actual theoretical gas temperature T_c and the next smallest and smallest curves are the response curves of junctions having diameter ratios of 1:2 respectively. These response curves were then fed into a computer program which calculated the gradients at each point and then found the corrected temperatures for both curves. The results, seen in Fig. (2), showed that the corrected temperatures for both response curves were identical and that they matched very closely (to within several decimal places) the actual gas temperature. Moreover, it was found that the time constant, α , could be calculated using values only for the response gradients, the response difference $(T_{j2} - T_{j1})$, the diameter ratio, and the value C_2 . The physical properties of the junctions were not a factor as long as both junctions were of the same material. Similar calculations were performed for various diameter ratios giving similar positive results, thus

showing that this technique works well in theory for any given diameter ratio.

A second theoretical model was set up to simulate an actual experiment in which two thermocouples of differing diameters at some initial temperature T_1 were suddenly exposed to a constant temperature environment T_2 . The response for each junction to such a step change looks like:

$$T_j(t) = T_2 - (T_1 - T_2) e^{-(t/\alpha)}$$

In order to match the experimental set up as closely as possible the diameter ratio was picked as 1:2 and T_1 and T_2 were chosen as 0.2 and 2.5 respectively (T 's are in volts). Again, the compensation was accurate when using a diameter ratio of 1:2. This time, however, it was decided to test the sensitivity of the compensation technique to the diameter ratio by trying to compensate with ratios somewhat different than that of 1:2. This was done and it was found that there was a great deal of sensitivity to this factor. Being off by even 1% (eg 0.99:2) from the actual ratio produced dramatically different results. Fig. (3) shows compensation curves for various diameter ratios.

It should be noted that a pattern was seen in the compensation curves as the diameter ratio was changed. When the guessed ratio began below that of the correct ratio (eg 0.1:2) and was increased, the compensation curve would slowly increase and approach the correct gas temperature. When the guessed ratio equalled that of the correct ratio, the compensation curve gave the actual gas temperature T_2 . When the guessed ratio surpassed the correct ratio (eg 3:2), the compensation curve would jump from positive to negative values in such a fashion that it began to approach the lower response curve. All this suggests a means of calibrating a probe to the correct diameter ratio by looking for some sort of convergence on an established gas temperature.

Sensitivity of this method to fluctuations in the response signal was also examined. This was done by adding a small amount of random noise to the theoretical response curves. The noise was on the order of 0.1% to 0.01% of the magnitude of the signal. Even by incorporating such small deviations into the responses, the compensation was greatly affected. Relatively large fluctuations in the compensated temperature could be brought on by only

minor fluctuations in the response curves. This suggests that in an experimental situation the signal should have a high signal to noise ratio for correct compensation to occur.

EXPERIMENTAL

The test configuration set up to verify this compensation technique involved placing a dual-junction probe in an environment of pulsed hot air of known temperature in order to provide a series of response curves which could be used to find the instantaneous gas temperature. Hot air was provided by passing compressed air through an electrical resistance heater which typically supplied values between 400°F and 600°F. The hot flow was then broken up using a rotating steel plate with a hole drilled through it. The test probe and the air supply were then placed on either side of the spinning chopper wheel and a laser and PMT were used to mark the onset of exposure to the hot air each time the hole passed in front of the probe. Typical wheel speeds, measured with a strobe, ranged between 10 Hz and 50 Hz.

Data was collected using a 386-based computer containing a DT2831-G A/D board with a maximum throughput rate of 250 kHz. A minimum of three channels were required for this experiment; one for each thermocouple and one for the laser. In future, the laser input channel may be replaced by an LDV input signal. Data was collected at a rate of 10 kHz and was amplified at a gain of 1000 and low pass filtered with the filter setting on 10 kHz.

The thermocouple wires were made of Pt/Pt-13%Rh and were of 1, 2 or 3 mil diameter, whereas the lead wire was 15 mil diameter. Probes using 1/2 mil and 1/3 mil thermocouples were constructed. The junctions were flame welded using a small hydrogen torch that created a bead which was typically four times the wire diameter and was usually not a perfect sphere. The smallest beads that most resembled spheres were chosen for the construction of each probe. The junction wires were connected to the lead wires using an electrical welder. This proved, through trial and error, to give sufficient strength to withstand flows of approximately 150 ft/s. An attempt

was made to place the junctions as close together as possible, but a spatial separation of several millimeters often could not be helped.

The probes used to test flowfield effects did not have actual thermocouple junctions but were merely the outer casings bent into the two configurations shown in Fig. (4). The probe bent similar to a pitot probe was rotated 90° in the flowfield and tested both ways. The probes were tested in a simple dump combustor with the LDV focused approximately 1-2 mm from the end of the probe.

RESULTS

Compensation was first attempted using data from a test with a 1 and 2 mil wire probe where the wheel speed was 10 Hz and the measured steady-state gas temperature was around 480°F (1.8 mV). The measured bead sizes were 4.3 mil and 8.0 mil in diameter. Feeding the information from this test straight into the compensation algorithm gave poor results. The computed gas temperatures varied widely and did not appear to approach the measured temperature of the hot air flow. In light of the theoretical model, however, this can be explained for two reasons. First, the measured response curves were not noise free and contained numerous small fluctuations. The affects of these fluctuations are, as previously mentioned, very important when taking point-by-point derivatives and can cause large variations in the computed results. Second, the measured diameters give a ratio that may not be the true ratio between the two junctions. Since the diameter ratio is really a ratio between the two masses, there may be some sort of "effective" ratio which cannot be determined in such a straightforward manner as optical measurement.

In light of these results, an ensemble average of fifty responses was made in order to help eliminate the effects of noise. This averaged curve is shown in Fig. (5). Note that the signals do not reach the same steady state temperature. This may be due to the spatial differences between the two junctions or possibly to the difference in the radiative cooling of each bead as it sits in still, ambient air between each hot pulse. The effects of this difference are as yet unclear, and so another data set was created in which

one of the curves was time shifted in order that both had the same steady state value and began to rise at the same time (see Fig. (6)).

Compensation was carried out on both the shifted and non-shifted curves in such a way as to find the effective diameter ratio. Various ratios were tried until it appeared that the compensated values began to converge on some constant value. In the 10 Hz case, the corrected gas values seemed to converge near 1.5 mV to 2.0 mV at a ratio of 2.75:8. In this way, the probe's effective diameter ratio had been calibrated to some degree of accuracy. This ratio was then used to compensate for the ensemble averaged curve at a wheel speed of 20 Hz. The results again appeared to be somewhat favorable, giving values within a range relatively close to that of the measured gas temperature (1.8 mv). However, enough noise still remained in the signals of both the 10 Hz and 20 Hz cases to cause the corrected values to fluctuate to a degree that made it very difficult to identify at what ratio the numbers actually converged. A polynomial curve fit was made to try to eliminate the remaining noise, but the particular fit that was a poor one. Some sort of curve fit must be made, however, in order to keep the corrected values from varying so wildly. Enough similarity did exist between the experiment and the model to indicate that this technique has definite potential.

The results of the LDV experiment were largely unsurprising. There were no great differences in the flow effects of the three probe configurations examined in the simple dump combustor. Future experiments may examine the effects in a swirl combustor in which a greater difference may arise.

CONCLUSION

Based on theoretical modelling, temperature compensation in the manner described appears to be possible. However, due to the inherent sensitivity of the technique on the signal to noise ratio and the ratios of the junction diameters, experimental implementation of this method is not so simple. Since the effective diameter ratio is not directly attainable, it appears that a calibration routine is necessary to experimentally determine this factor. This can only be accomplished if the signal is essentially free of

noise, and the best way to do this appears to be a combination of ensemble averaging and curve fitting.

Clearly, more work needs to be done regarding the method of cleaning up the signal and being able to converge on the correct diameter ratio. This may also involve a better understanding of the fluid mechanics involved in performing the described experiment.

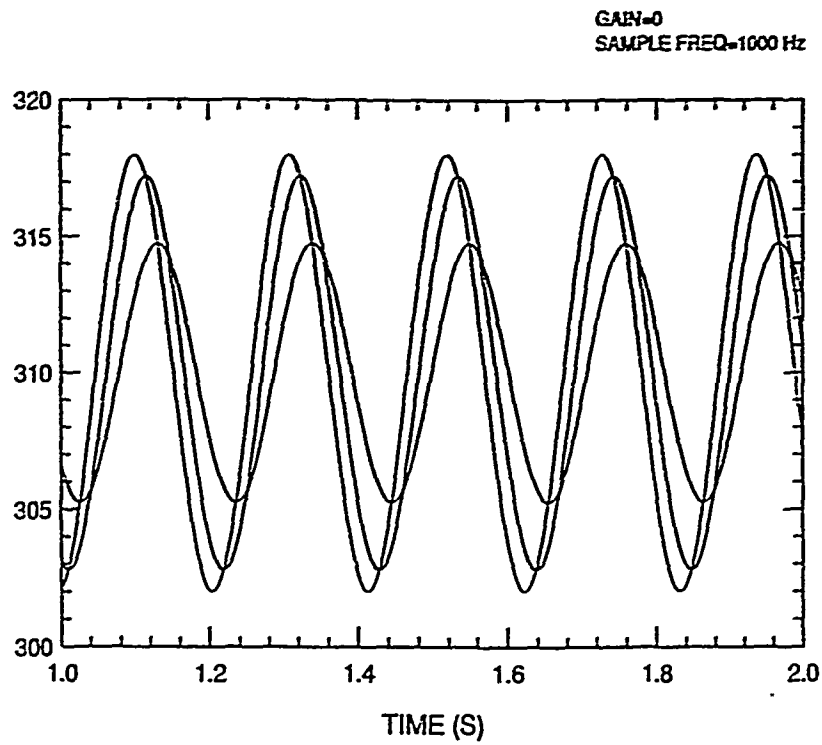


Figure 1. Sinusoidal Gas Temperature With Corresponding Response Signals

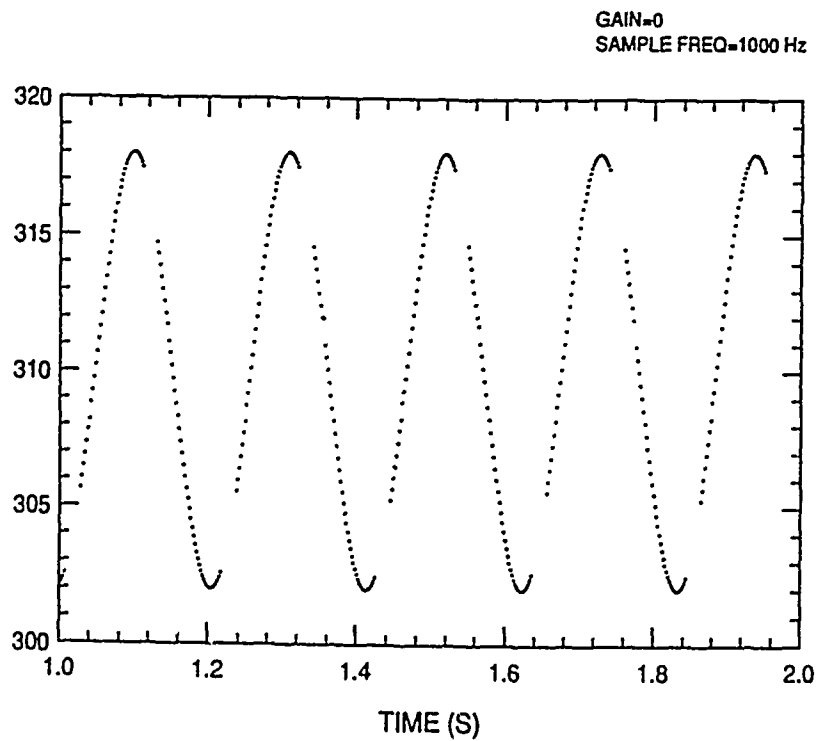


Figure 2. Compensated Gas Temperatures Based On Sinusoidal Model

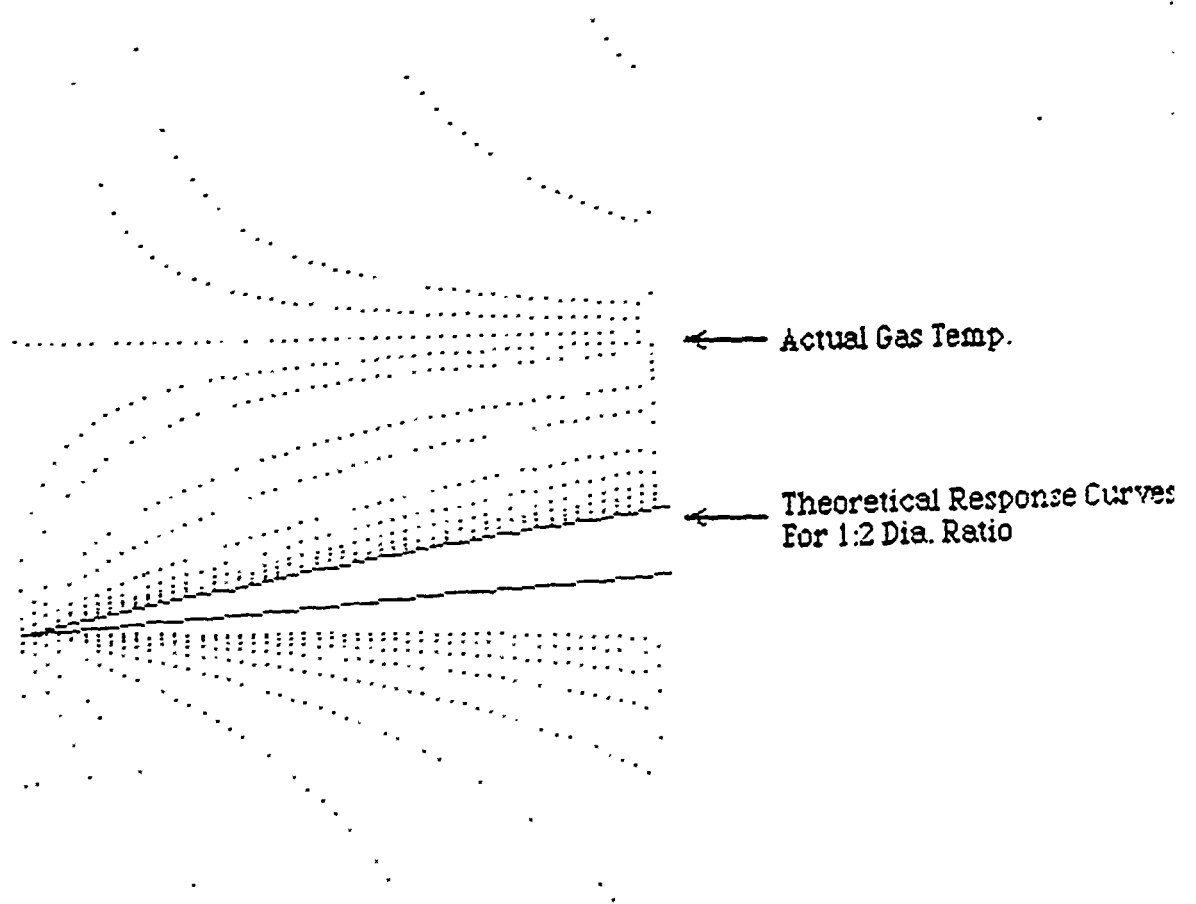


Figure 3. Effect Of Diameter Ratio On Compensation

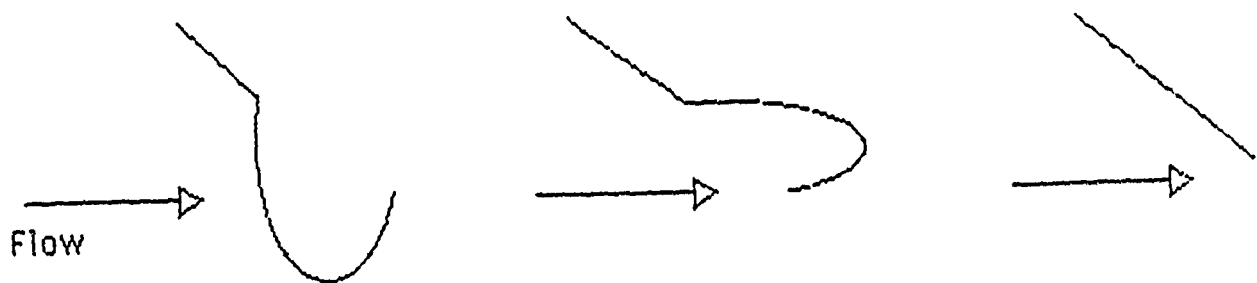


Figure 4. Various Probe Geometries Tested

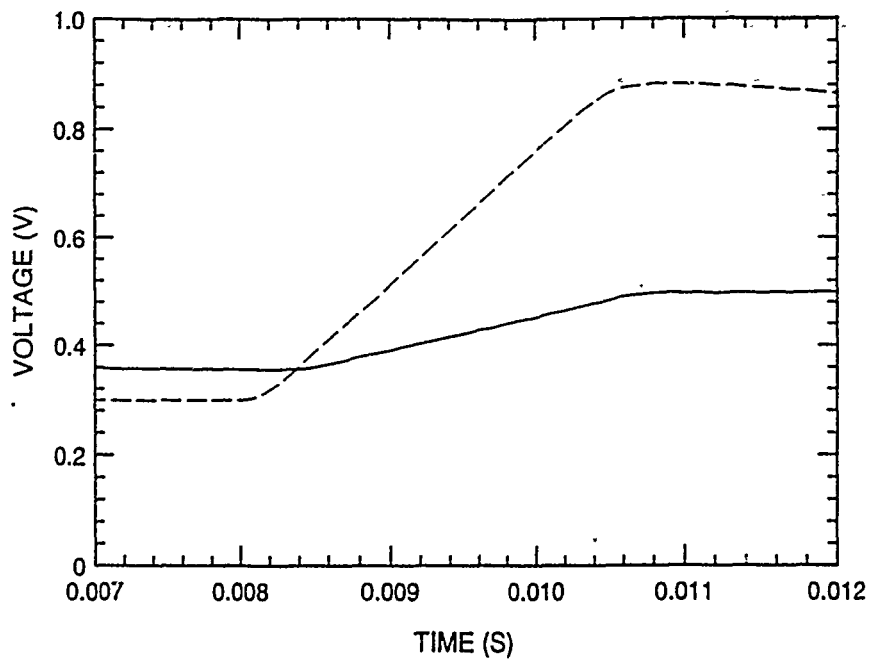


Figure 5. Ensemble Averaged Response At A Wheel Speed Of 10 Hz

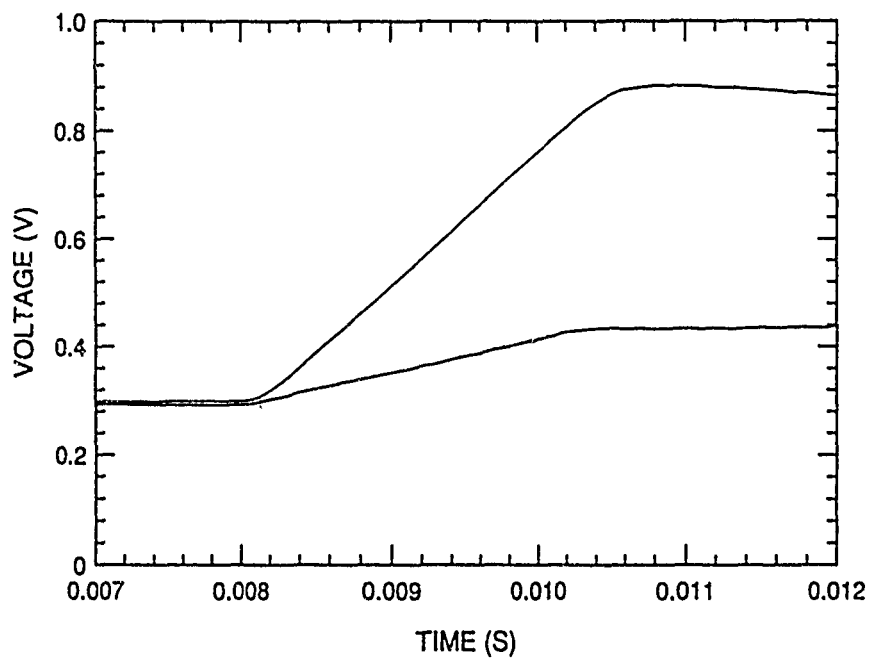


Figure 6. Time-Shifted Average Responses

1991 USAF-RDL, INC. SUMMER FACULTY AND GRADUATE STUDENT
RESEARCH PROGRAM/

Sponsored by the
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the
Research Development Laboratories, Inc.

FINAL REPORT

Investigation of the Combustion Characteristics of Swirled Injectors in a
Confined Coannular System with a Sudden Expansion

Prepared by: Paul O. Hedman, Professor
David L. Warren, Master Candidate

Departments and Chemical and Mechanical Engineering

University: Brigham Young University
Provo, Utah 64602

Research Location: Aero Propulsion and Power Laboratory
Wright-Patterson AFB, OH 45433

USAF Researcher: W. M. Roquemore, Ph.D.

Date: September 11, 1991

ABSTRACT

This report contains a brief summary of the work done to investigate the operational characteristics of a burner that was designed to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor." The burner, referred to as the Pratt & Whitney Task 150 Combustor, uses a swirling fuel injector from an actual Pratt & Whitney turbojet engine installed in a sudden expansion combustor that closely simulates the geometry of a combustor from an actual jet engine. The Task 150 configuration has been configured so that the geometry around the injector is nearly axi-symmetric, but the combustor incorporates quartz windows so that optical (laser based) instruments can be used to make measurements in the flame. The Task 150 configuration uses a swirling injector similar to those used in the Task 200 combustor, and the inlet diffuser sections and inconel chimney of the Pratt & Whitney Task 100 burner. This unique configuration allows complex diagnostic measurements to be measured in a simpler geometry than the Task 200 combustor, but embodies most of the features of an actual jet engine combustor in an axi-symmetric configuration that is easier to mathematically model.

The primary effort during this summers AFOSR sponsored research program for faculty and graduate students was to assemble the hardware, and investigate the basic operational behavior of the burner. It was found that the flame would exhibit very different operating characteristics over the operating stoichiometric range. The flame would be attached to the burner or lifted from the burner as the fuel equivalence ratio was changed. Film images and video tape of the various operating modes as a function of fuel equivalence ratio were obtained.

Measurements of the fuel equivalence ratio at lean blow out as a function of air flow rate were also obtained. At high air flow rates, the flame would blow out for both injectors tested from a well lifted flame that was being stabilized on a downstream recirculation zone. At low air flow rates, the low swirl injector would still blow out from the flame stabilized on the down stream recirculation zone. However, at low air flow rates, the flame on the high swirl nozzle would remain attached to the center core of the fuel injector right up to lean blow out. As a consequence, the fuel equivalence ratios (based on total fuel and total air flow) where lean blow out occurred with the high swirl nozzle were very much lower (ca 0.3) than that observed for the low swirl injector or the high swirl nozzle at high air flow rates (ca 0.5).

The pressure drop across the low swirl and high swirl injectors were determined as a function of total air flow rate through the combustor. In a separate investigation, the partitioning of the flows through the dome jets, the insert jets, the primary air swirler and the secondary air swirler was determined. The effect of fuel flow on the pressure drop across the injectors and its effect on the flow partitioning was also investigated by introducing CO₂ into the fuel passage at varying flow rates. How the air flows partition between the various flow passages in an injector is expected to have a major impact on how well that injector operates.

Measurements of wall pressures and wall temperatures were made in an attempt to better understand the locations of the various flame zones in the combustor. These measurements have indicated that the location of the recirculation zones moves very little as operating condition is changed. They also indicated a second recirculation zone in the dome region that is caused by the dome cooling jets.

I. INTRODUCTION

This report presents a brief summary of results of an initial investigation to determine the flame characteristics when swirling fuel injectors from actual Pratt-Whitney jet engines are installed in a burner with a sudden expansion (Pratt and Whitney Task 150 Combustor). This work provides a bridge between the work where the combustion characteristics of confined, coannular fuel and air jets are discharged into a sudden expansion (Pratt and Whitney Task 100 Combustor), and the Task 200 combustor where four swirling fuel injectors from actual Pratt-Whitney jet engines are installed in a a rectangular combustion chamber that simulates a segment of a real jet engine combustor. The advantage of the Pratt and Whitney Task 150 Combustor is that it allows the actual combustion characteristics of a real injector to be

investigated in a simpler geometry where various diagnostic measurements (primarily laser based optical measurements) can be more easily made. The Task 100 and 150 combustor configurations have been specifically developed to study the phenomenon of lean blowout (LBO) in modern annular aircraft gas turbine combustors. The combustor has been carefully designed (Sturgess, et al. 1990) to "specifically reproduce recirculation patterns and LBO processes that occur in a real gas turbine combustor."

The Task 100 combustor consists of coaxial jets with a 29 mm diameter central fuel jet surrounded by a 40 mm diameter annular air jet. The jets are located in the center of a 150 mm diameter duct. A sudden expansion, rearward facing bluff body, with a step height of 55 mm, is located at the exit plane of the coaxial jets. The combustor test section incorporates flat quartz windows to accommodate laser and other optical access, but uses a metal shell with metal corner fillets to reduce the vorticity concentration and eliminate its effect of the bulk flowfield in the combustor. This box-section combustor with corner fillets allows reasonable optical access, while providing a cross section that approximates a two-dimensional axisymmetric cross section. The bluff body provides a recirculation region that can stabilize the flame.

The Task 150 combustor configuration utilizes the basic Task 100 hardware, but replaces the confined, coannular jets with an insert and an actual swirling fuel injector from a Pratt and Whitney jet engine. A schematic drawing of the Task 150 Combustor is shown in Figure 1. A drawing that shows the installation of the fuel injector in greater detail is presented in Figure 2. Two different fuel injectors were used for this study, a high swirl injector, and a low swirl injector. The high swirl injector is referred to by representatives of Pratt and Whitney as a "bill of materials injector" used in production engines. The low swirl injector was reported to match the characteristics of the injectors supplied by Pratt and Whitney for use in the Task 200 combustor.

The objective of the project was to determine the combustion and flow characteristics of the Task 150 burner over a range of operating conditions. Specifically, the study was to characterize the broad operating characteristics of the Task 150 burner with both high and low swirl injectors. The initial characterization included the following experimental work: flow meter calibration, checkout experiments, flame characterization experiments which were recorded on film and video tape, lean blow out measurements, determination of the flow partitioning between the injector passages as a function of air flow rate, and measurements of wall differential pressure and wall temperature measurements.

The results of this study relate to a flame blowout modeling study being conducted by other investigators from Pratt and Whitney (Sturgess, et al. 1990). A secondary purpose of this study has been to collect data that could be used in validating a computer code that predicts the flame phenomena and blowout limits (Sturgess, et al. 1990).

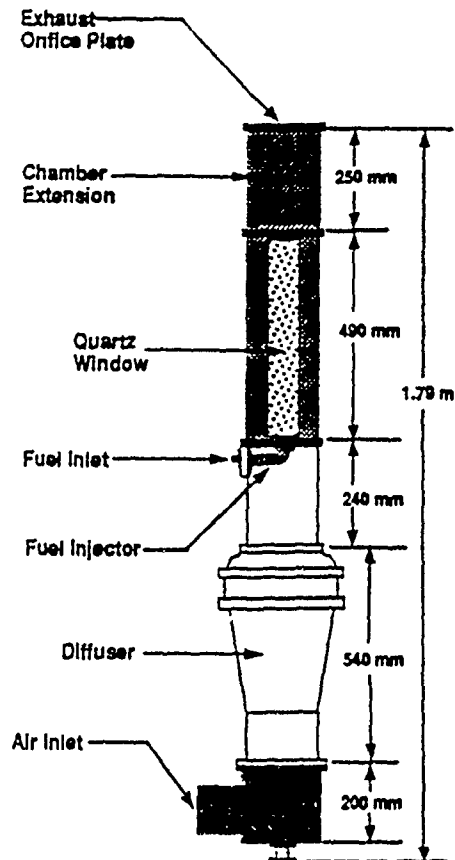


Figure 1. Schematic of the Task 150
Pratt and Whitney Combustor

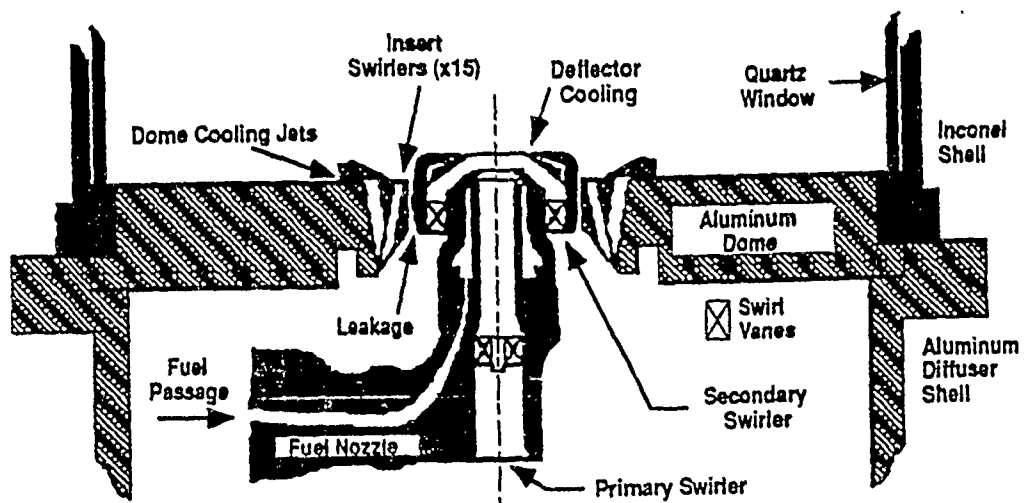


Figure 2 - Task 150 Burner Configuration

The page constraints of this report prevent a detailed summary and discussion of all of the experimental results obtained. Consequently, only example results are presented. A more complete compilation of the data has been prepared in the form of an appendix which has been supplied to the Air Force sponsor at Wright Patterson Air Force Base.

II. RESULTS

Calibration of Flow Meters

The main instruments used in these test series were four flow meters. These four included one for the air (6000 slpm air capacity), another for the nitrogen (1000 slpm air capacity), and two (300 and 45 slpm air capacity) for the fuel. The largest reference standard available was a laminar flow cell with approximately a 300 slpm limitation. This arrangement worked satisfactorily for the fuel meters, and showed the nitrogen meter to be linear up to the flow cell's limit. However, the air meter, not being designed to run at this low of flow rate, showed serious but predictable inaccuracies. Careful calibration against the laminar flow cell at low flow rates and against the linear nitrogen flow meter at higher air flow rates allowed the air flow meter to be calibrated over the flow rate range of interest. Figure 3 presents the final calibration of the air flow meter used in this study.

The calibrations of the fuel flowmeters were found to be very close to linear, and to match the factory calibrations very closely. The calibration equations obtained for the two fuel meters used are shown below.

0 to 300 slpm air (0 to 108 slpm propane) meter:

$$Y \text{ (actual flow)} = 0.99158 + 0.988 \cdot X \text{ (indicated flow);} \quad R^2 = 1.000$$

0 to 45 slpm air (0 to 16.2 slpm propane) meter:

$$Y \text{ (actual flow)} = -0.018719 + 0.938 \cdot X \text{ (indicated flow);} \quad R^2 = 0.999$$

Flame Characterization

One of the traits of the Task 150 burner is the wide variety of flame structures that are observed. Each flame shape indicates a different mode of operation, which differ from one another in the location of the flame fronts, or by some structure such as thickness or intensity. These different structures might arise from changes in the flow fields, mixing patterns, or fuel equivalence ratio. The differences and similarities of the flame structures for both nozzles together with the Task 100 configuration may provide significant

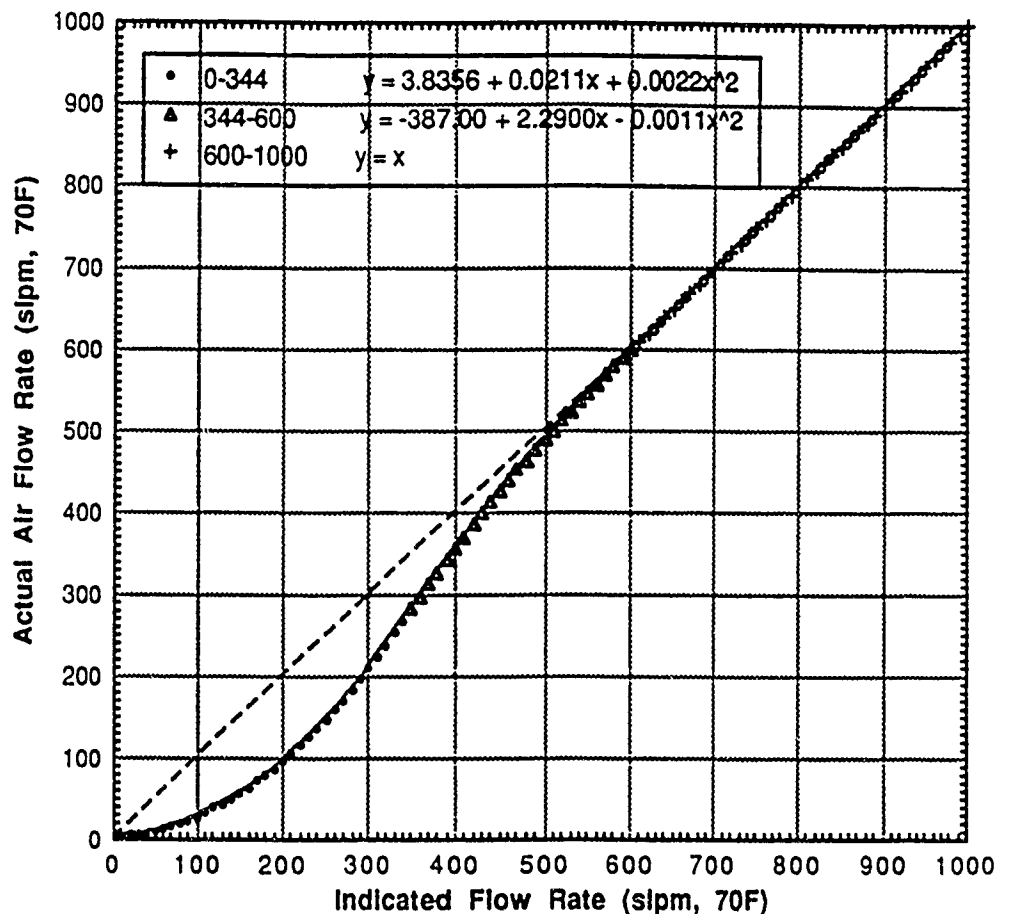


Figure 3 - Final Air Flow Meter Flow Calibration

insights to the combustion processes. Characterization simply defines the modes and the differences that differentiate the structures.

The shape of the flame, at the minimum, provides qualitative information on the mixing process and location of flame fronts. Such information can yield precious insights into what processes are present. Flames fronts exist because fuel and oxidizer have been transported to a point where combustion can be supported. The location of these fronts relative to the outlet orifices of fuel and air are of obvious interest.

Flame structure is also important in the macroscopic effects it can have. In the high swirl burner near lean blow out, for example, the flame is very small, with much of the air bypassing the actual combustion region. This small flame structure supported much lower fuel equivalence ratios than the comparable low swirl injector, which had a much larger flame structure to maintain. These structures not only have an effect in

laboratory experiments but also affect the real-life operation of a turbojet engine. Understanding of the causes and effects of flame structures may be of significant help to the injector designer.

The fuel equivalence ratios where the transitions from one flame structure to another were determined as a function of air flow rate. The flames for both injectors were attached to the outside of the insert air jets when the burner was operated very fuel rich. The flame would then lift, reattach, and lift again as the fuel equivalence ratio was progressively reduced depending on the injector (high swirl versus low swirl) and the air flow rate. During the reattachment phase, the flame would take on many of the characteristics of a tornado. Consequently, the terminology of funnel cloud, tornado, and debris cloud were adopted to describe some of the observed flame structure.

The flame characterization of the Task 150 combustor was in large part carried out by visual observations. Still film photographs were taken of the different structures. These images were digitized and manipulated using various computer programs into the isochromatic contour plots found in Figures 4 and 5 for the low swirl and high swirl injectors respectively. As can be seen, the flame for both nozzles look alike at very fuel rich conditions. The flame is attached to the insert jets, in a manner similar to the Task 100 burner at rich conditions. Unlike the Task 100, these flames are very short, presumably because the oxygen in the air was being depleted much faster due to the swirling motion mixing the gases. As the amount of fuel was reduced, the flames lifted and stabilized on some downstream recirculation zone that appeared to be associated with the injector. The primary combustion zone lengthened as relatively more oxygen entered the chamber. At this point the two nozzles diverged in flame characteristics. The low swirl nozzle would abruptly switch to an attached cone flame. Although this cone seemed similar in appearance to the high swirl flame near lean blow out, it was much different. The attached cone was much more vigorous, and produced a howling roar, giving no indication of being near blow out. As fuel was further reduced, the attached cone would once again lift and stabilize on a downstream recirculation zone associated with the injector. At this point, however, a flame structure could still be perceived, as illustrated in Figure 4. As the fuel equivalence ratio continued to drop, the cone dissipated and became a shapeless, separated flame that appeared to be stabilized on the larger recirculation zone associated with the chamber itself. It is believed in this structure, that the nozzle no longer exerted an influence on the flame, and the flame was being stabilized on conditions sustained by the chamber itself. With the lean lifted flame structures, flames with the low swirl nozzle once again took on the appearance of flames seen in the Task 100 combustor.

With the high swirl nozzle, after the rich lifted condition, the swirling air began to stabilize a flame in the center of the combustor, with a result that behaved much like a tornado. As seen in Figure 5, a funnel cloud formed within the rich lifted flame and gradually descended as the fuel flow was decreased. Eventually, the funnel appeared to disappear into the inside of the primary swirler passage in the injector.

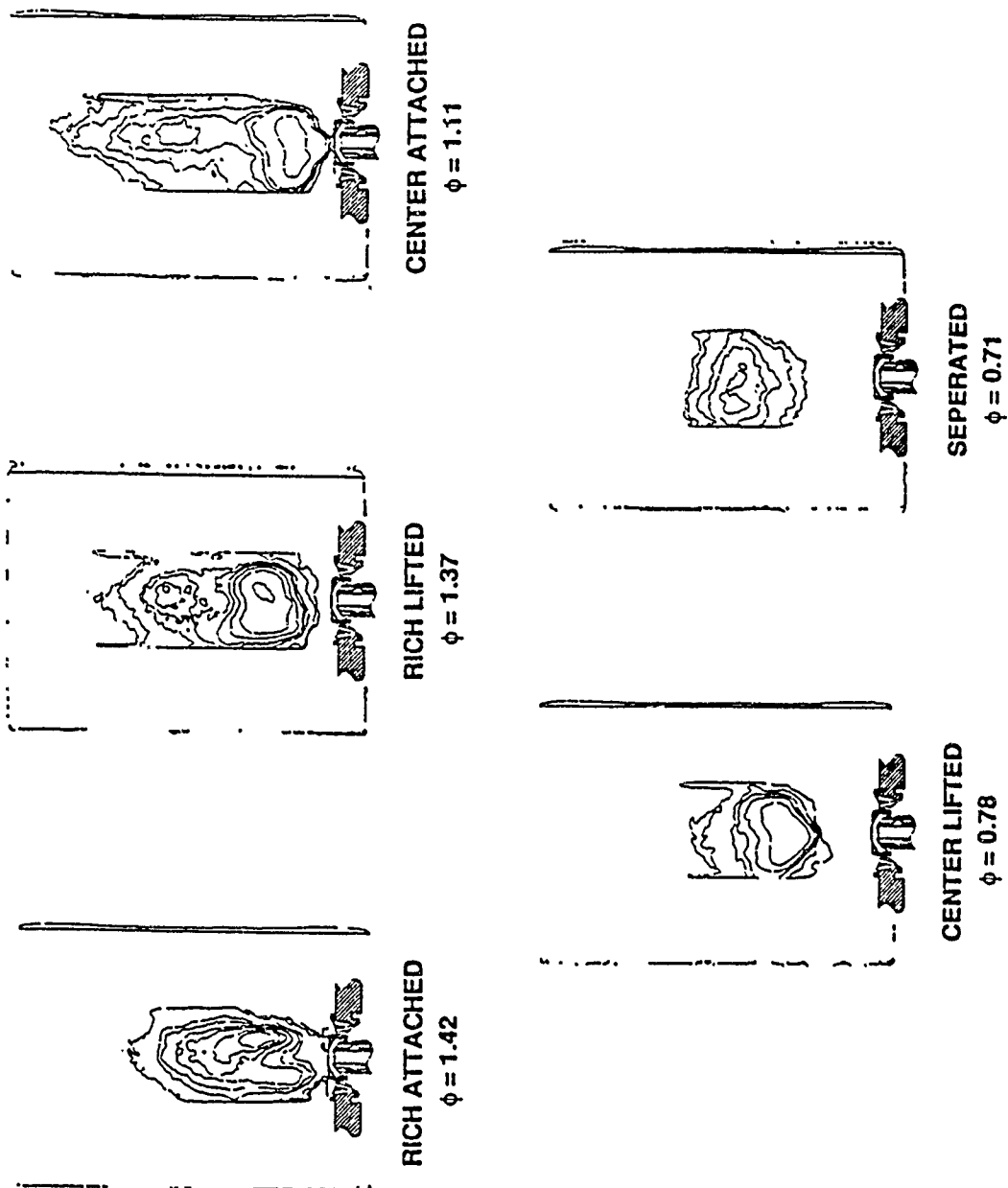


Figure 4. The five different flame structures observed with the low swirl nozzle. (Air flow = 310 slpm, 70 F).

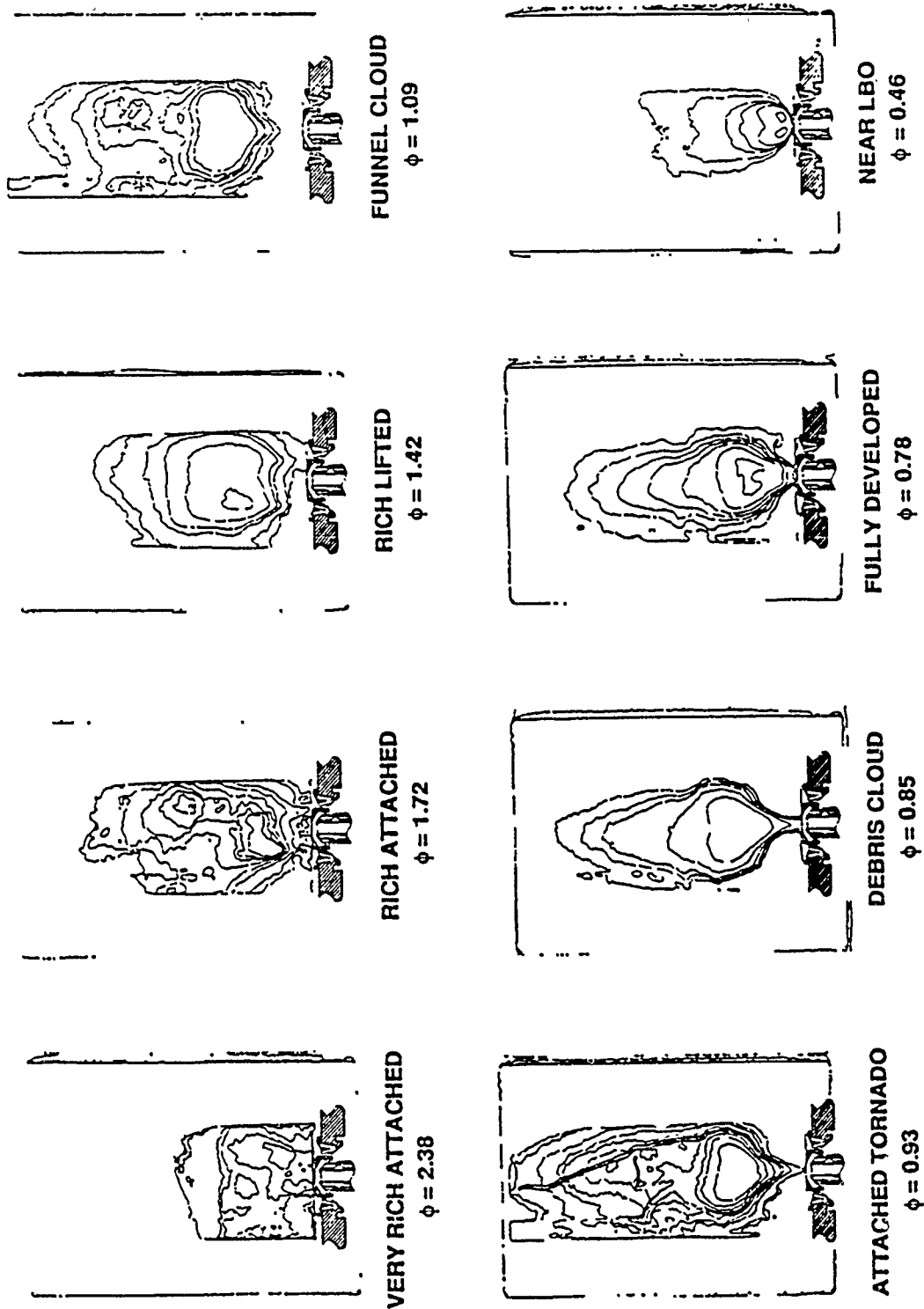


Figure 5. The eight different flame structures observed with the high swirl nozzle. (Air flow = 310 slpm, 70 F)

A flame in the shape of a little bowl, which looked much like the debris cloud of a tornado, attached to the nozzle, formed on the outside of the tornado cloud. This detail was lost in the process of converting from still photograph to Figure 5. Continued reduction of the fuel flow resulted in the growth in size and intensity of this cloud while simultaneously decreasing the funnel cloud. The disappearance of the funnel cloud marked a transition to a fully developed flame that was strongly attached to the center of the injector. At low air flow rates, this strongly attached flame would weaken until the lean blow out limit was reached. At high air flow rates, the flame would once again lift, attach to down stream recirculation zones, and eventually blow out from the separated flame structure, much as did the Task 100 Combustor, and the LS Task 150 Combustor.

One of the major factors in flame structure was fuel equivalence ratio. At a given air flow, the fuel would be reduced until a transition was judged to have been reached. These observations were not easily made. With the low swirl nozzle, most transitions between modes were quite abrupt, but the flame would oscillate from one mode to another (without obvious cause) over a narrow range of fuel equivalence ratios. In these cases, the practice was to wait until a complete (no jumping between modes) transition was reached. With the high swirl nozzle, however, the behavior was quite different. Instead of abrupt changes, the flame flowed smoothly from one mode to another. While eliminating the problem of flicker, these smooth transitions also left no sharp break point in flame behavior. Easily recognizable events, such as when the tornado appeared to touch the nozzle, were chosen to differentiate operating modes. Even still, events such as the onset of the funnel cloud were in large part a matter of judgment. With these understandings, Figures 6 and 7 quantify the different fuel equivalence ratio regimes in which the various modes operate.

Lean Blow Out Experiments

Lean blow out tests were conducted at various times during the course of the summer research effort. These measurements were conducted with the high swirl and low swirl nozzles, but were conducted with the 10 inch extension, and the 45% exhaust orifice plate installed. A more systematic set of measurements was made as part of the tests to measure wall pressure and temperature with various operating configurations, and over a range of both air and nitrogen flow rates. While making these tests, the burner was operated near LBO while the wall pressure and temperature measurements were made. Once this data had been collected, the burner would be further leaned until lean blow out was achieved and recorded. Frequently, the burner would be relighted, and the lean blow out measurement repeated.

It is beyond the scope of this document to report all of the lean blow out data taken during the course of the summer research program. The additional data is available on request. An example set of representative blow out data is contained in Figure 8 for both the low swirl and high swirl nozzles. For reasons not yet fully understood, the fuel equivalence ratios at lean blow out for air flow rates below about

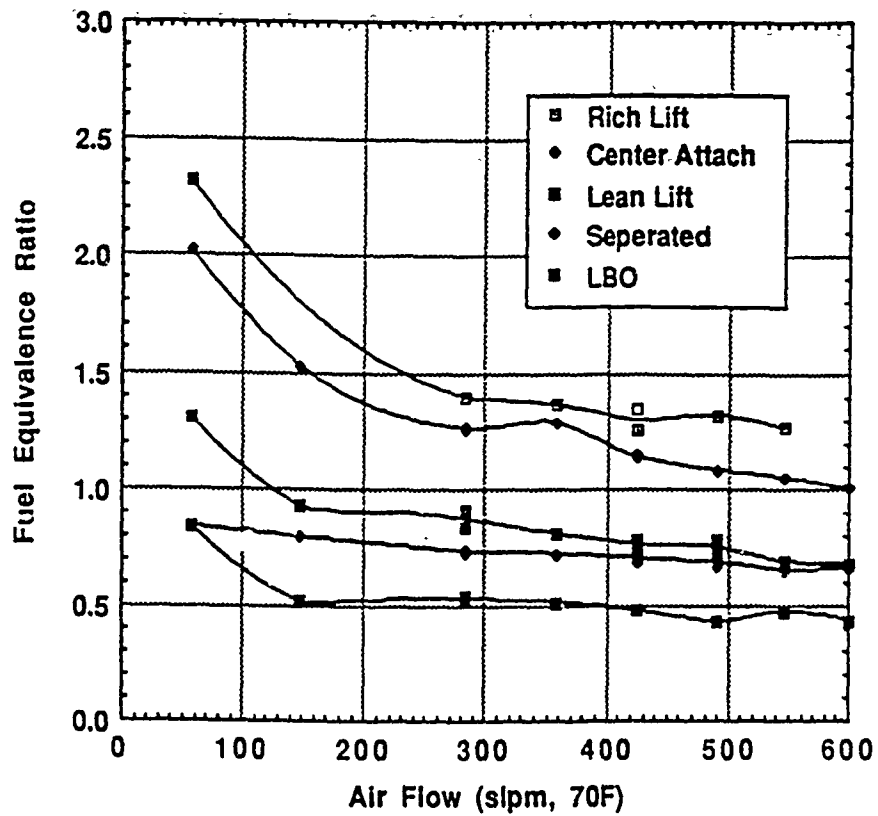


Figure 6 - Operating modes of the Task 150 Combustor with the Low Swirl Nozzle.

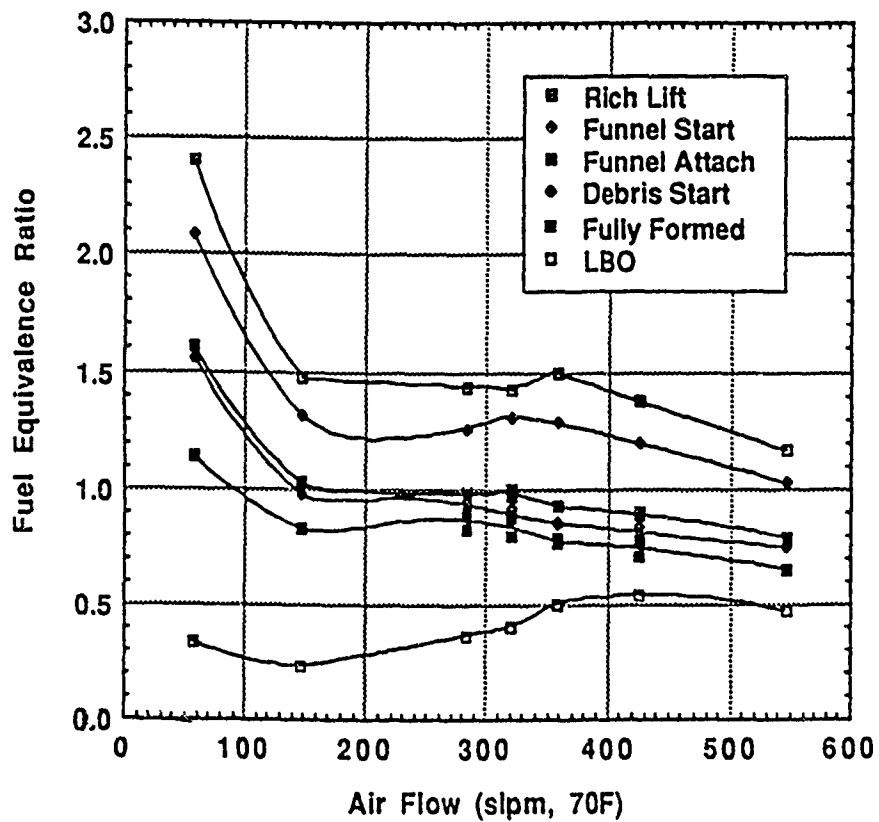


Figure 7 . Operating modes of the Task 150 Combustor with the High Swirl Nozzle.

400 slpm depart from the expected value of about 0.5. As Figure 8 shows, the ϕ at LBO for the low swirl fuel injector increases significantly as air flow rate is reduced below about 400 slpm (70 F). Conversely,

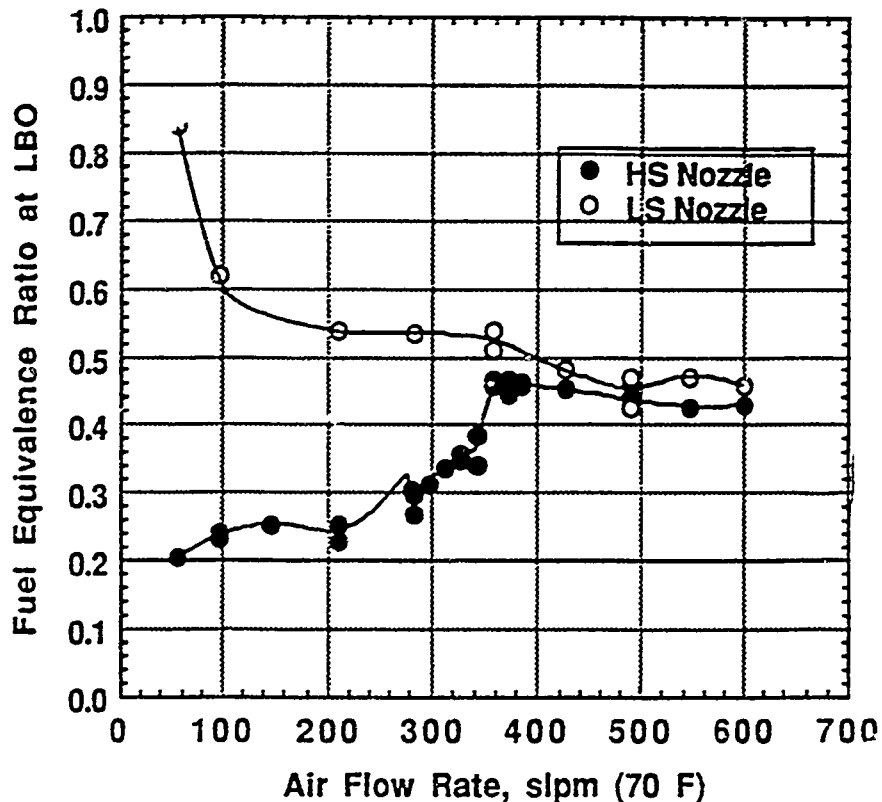


Figure 8 - Example Set of Lean Blow Out Data for Task 150 Combustor (10 inch Extension, 15% Orifice)

the ϕ at LBO for the high swirl fuel injector decreases markedly as air flow rate is reduced below about 800 slpm (70 F).

Time has not permitted a complete evaluation of this lean blow out data, but a few observations have been made. At high air flow rates, the flame just prior to blow out for both injectors tested was a well lifted flame that was being stabilized on a downstream recirculation zone. At low air flow rates, the low swirl injector would still blow out from the flame stabilized on the down stream recirculation zone. However, at low air flow rates (less than ca 600 slpm), the flame with the high swirl nozzle would remain attached to the center core of the fuel injector right up to lean blow out. As a consequence, the fuel equivalence ratios (based on total fuel and total air flow) where lean blow out occurred with the high swirl nozzle were very much lower (ca 0.4) than that observed for the low swirl injector or the high swirl nozzle at high air flow rates (ca 0.5).

It was evident from observations of the flame associated with the high swirl burner at the lower air flow rates, that the volume of the flame zone was very small compared to the volume of the combustion

chamber, and that some of the combustion air was bypassing the combustion zone. Thus analysis of the flame in terms of a "loading parameter" must be done very carefully.

It is suspected that the flame zone is still behaving like a well stirred reactor, but that the local stoichiometry is not well represented by the overall fuel equivalence ratio, the total air mass flow, and the total volume of the reactor. It is difficult to understand how a flame can be sustained at a fuel equivalence ratio well below the lean flammability limit ($f = 0.5$). It is suspected that the f at lean blow out would be much closer to 0.5 if the volume of the flame zone could be accurately estimated, and if some estimate of the actual mass of air and fuel entering the flame zone could be made.

Flow Splits

Measurements of pressure drop across the various flow passages in the fuel injectors installed in the Pratt & Whitney Task 150 combustor were used to estimate the partitioning of the flows between the primary swirler passage, the secondary swirl passage, the insert jets, and the dome jets. The various passages were plugged in sequence so that the flow rate through a single passage could be correlated with the pressure drop across the dome of the combustor.

Each set of air flow, pressure drop data was curve correlated using a quadratic least squares function. It was assumed that the flow correlations for the insert and dome jets would be the same for both the HS nozzle and the LS nozzle sets. In making the measurements through the insert and dome jets, the flow through the combined insert and dome jets was correlated with pressure, and the separate flow through the insert jets was correlated with pressure. The separate flow through the dome jets was then obtained by subtracting the insert flow correlation from the combined correlation. The separate equations for each of the four passages for both the HS and LS injectors were then used to determine the flow rate as a function of pressure drop for each of the passages, and the sum was used to determine the correlation between the total air flow rate and the pressure drop across the injector. With the total air flow determined, the percentage of flow through each injector passage was determined as a function of air flow rate. Figures 9 and 10 summarize the percentage of flows through the primary swirler passage, the secondary swirler passage, and the combined insert and dome jets for the high and low swirl nozzles respectively. The higher pressure drop through the HS primary flow passage and the attendant lower flow rate is apparent.

A question on the effect of fuel injection on the pressure drop across the fuel injector, and its effect on the partitioning of the flow between the various passages was raised near the end of this summer's faculty and graduate student research program. There was insufficient time to fully explore this effect. Nevertheless, a brief investigation into this effect was undertaken. CO₂ was introduced into the fuel orifice of both the HS and LS injectors. The insert and dome jets were completely blocked, since it was felt that fuel injection could not have any effect there. The flows in the primary swirler or the secondary

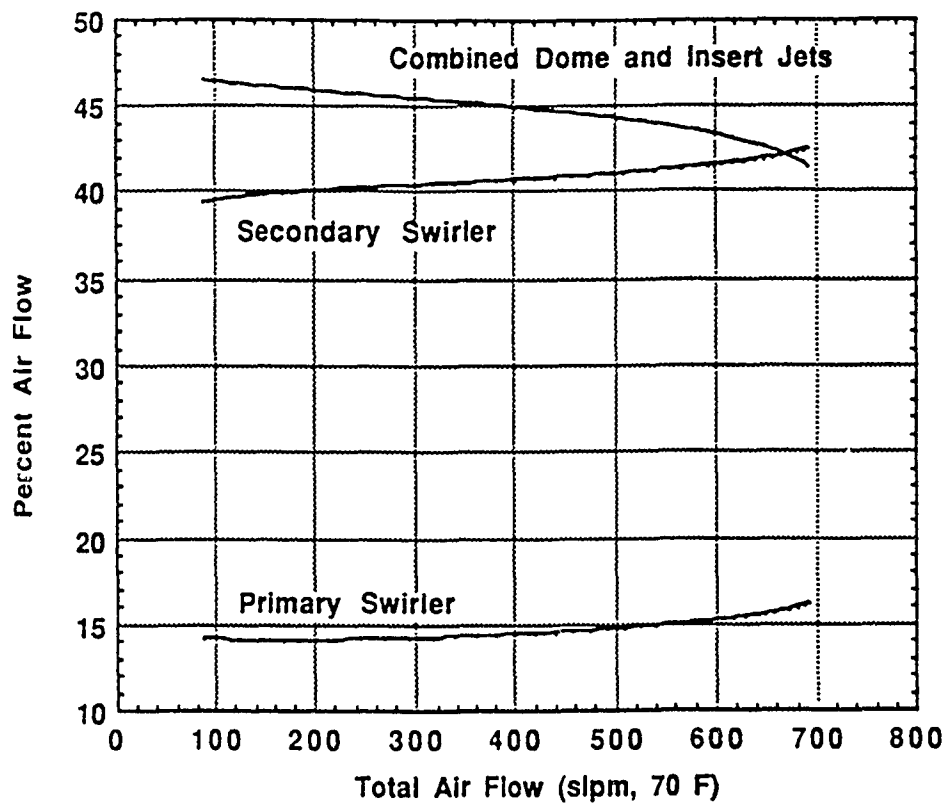


Figure 9 - Flow Splits for Task 150 Combustor with High Swirl Nozzle.

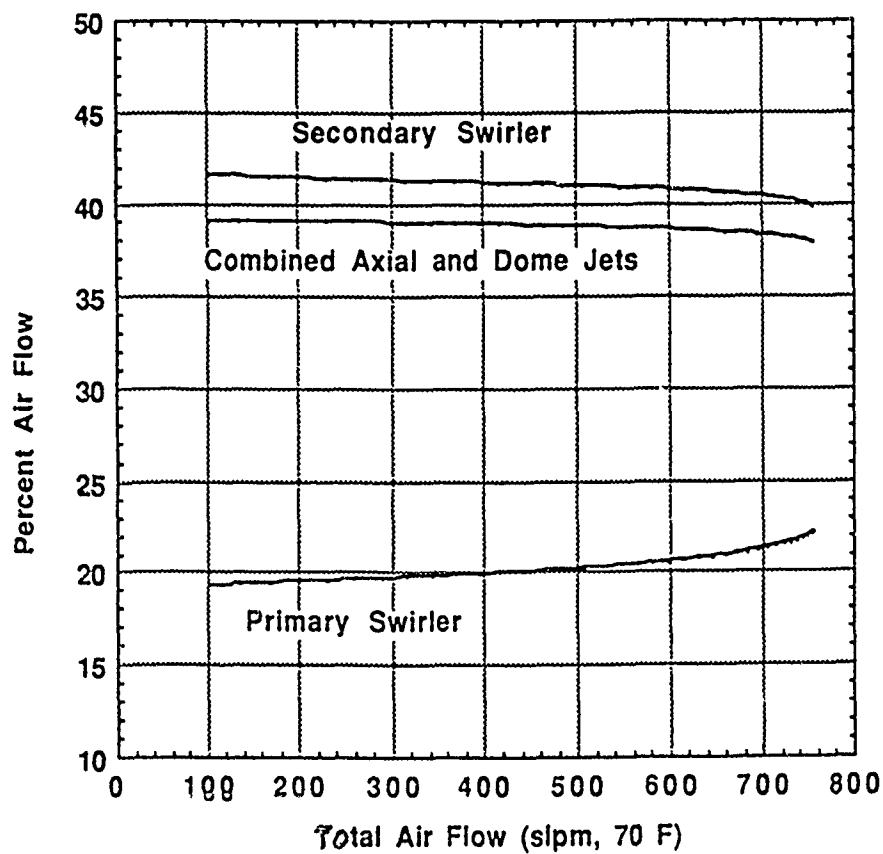


Figure 10 - Flow splits for Task 150 Combustor with Low Swirl Nozzle.

swirler were successively blocked so that the change in pressure across the injector could be monitored as a function of CO₂ flow rate through the fuel orifice. Four separate conditions were used, two with the HS nozzle, and two with the LS nozzle. The air flow rate required to achieve a given pressure drop (LS, 1.67 and 4.35 in. H₂O; and HS, 4.35 and 10.15 in H₂O) was set and held constant while the CO₂ flow rate was changed. The pressure change (referenced to the set pressure drop) as a function of CO₂ flow rate for the four cases is summarized in Figures 11 and 12. It can be seen that the introduction of CO₂ (fuel) has a very significant effect on the pressure drop across the passage, and by implication, on the flow rate through the passage for a given pressure drop. The effect is much more pronounced for the primary swirler passage than the secondary swirler passage. The effect can be either positive or negative depending on the injector (HS or LS) and air flow rate. That is, it can impede the flow causing an increase in the pressure drop, or it can act like an ejector causing a reduced pressure and consequent higher flow rate. Examination of the figures shows that the effect may be quite nonlinear and can vary from positive to negative back to positive depending on condition. It is evident that the addition of fuel will have a significant effect on the way that the flows split as they pass through the various passages of the injector. Further work in this area is warranted.

It should be noted that the percentage flow splits are thought to be quite accurate over the range of flow rates and delta pressure measured. This work needs to be extended to higher flow rates and higher delta pressures. Some of the actual combustion experiments conducted used air flow rates as high as 2500 slpm. This gave a delta pressure across the dome of the burner of about 4 psi. There were also some discrepancies when the total air flow rate as determined by the summation of the separate equations was compared to that measured directly with all passages open. There were some leaks in the facility that were identified at a later time that might have been affecting the results. Also, the flow through the very small gap between the injector and the dome insert may have been affecting the results. The introduction of CO₂ to simulate the effect of fuel injection was found, as was discussed above, to be significant and very nonlinear. It is recommended that additional work be undertaken to further investigate the flow partitioning through the various passages over a wider range of flow and pressure conditions, and with additional study of the effect of fuel addition on the flow splits.

Wall Pressure and Wall Temperature Measurements

Wall temperature and pressure measurements were taken with thermocouples and pressure taps located in wall plates at selected axial locations. Measurements were made at or near the lean flammability limit and for selected cases near the maximum heat release. Test configurations included combinations of the HS and LS nozzles, 0 in and 10 in chamber extensions, and 0%, 45%, and 62% exhaust orifice. Much of the data were collected without any supplemental nitrogen being added to the combustion air stream. There were some sets where a low levels (ca. 10%) and high levels (ca. 20%) of N₂ were added to the

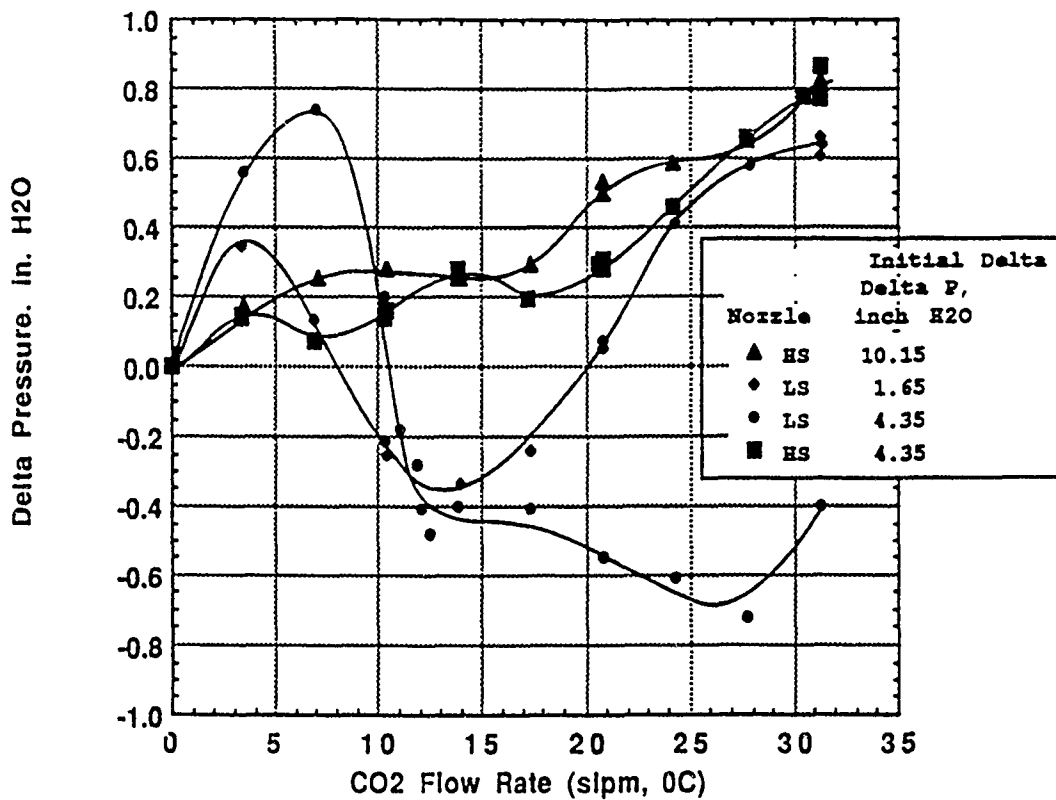


Figure 11 - Pressure Drop across Primary Swirler with Simulated Fuel (CO₂) Injection

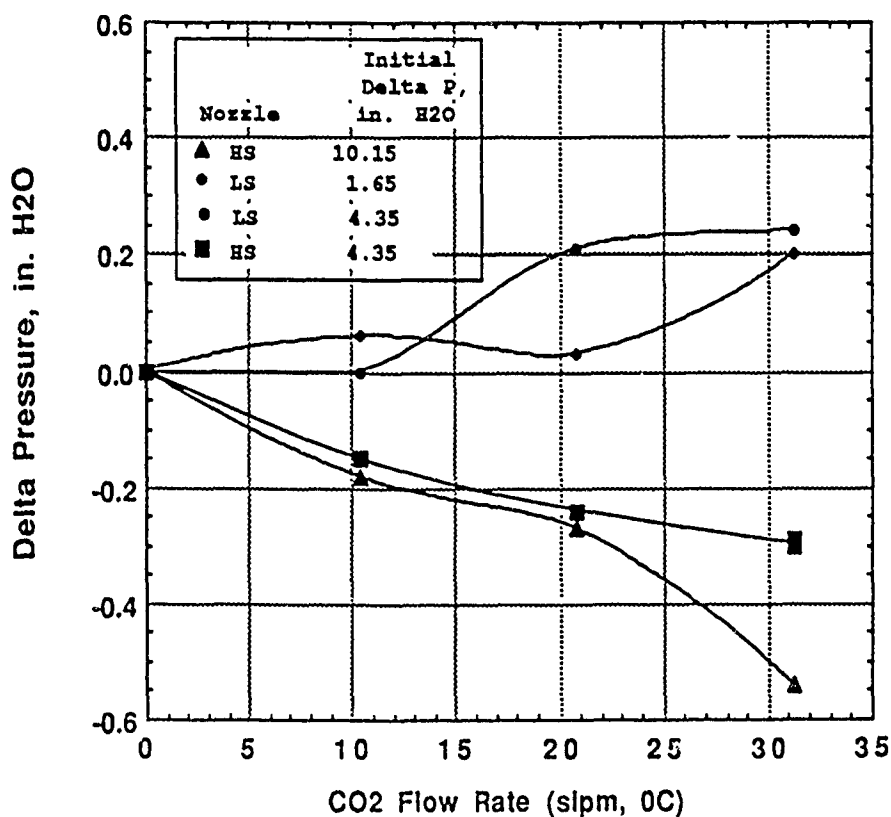


Figure 12 - Pressure Drop across Secondary Swirler with Simulated Fuel (CO₂) Injection

combustion air. In as much as time was limiting the latter test phases, A sparse test matrix was employed, i.e. only three levels of air flow, and zero or a high level of nitrogen were added to the combustion air.

With the burner exit open to the atmosphere, it was expected that the gage pressure at the wall would be indicative of the large-scale recirculation zone flow patterns. At points where the flow had a significant radial velocity, the pressure would change according to the stagnation pressure. In other regions, the pressure would be dominated by Bernoulli's equation--the higher the velocity, the lower the pressure. The most obvious result of the information collected was the invariability of the flow structure. This consistency can be seen in Figure 13, which shows all the data points taken. There were three zones in evidence. The first recirculation zone, occurring in the first 50 mm, was dominated by the dome jets. The second zone, located between 50 and 200 mm, consisted of a lower recirculation zone caused by the flow in the primary combustion volume. The last zone was the combustion products flowing through the chamber. A few data sets also showed evidence of another recirculation zone near the exit, set up by the orifice blockage plate. This structure remained invariant for a wide range of test conditions. The main effect of the extension and orifice plate was to shift the curve up or down (Figure 13). The air flow changed only the magnitudes of the pressures, but did not appreciably change the location of the zones. The principle nozzle effect was on the strength of the first (dome) zone, with the high swirl nozzle creating a stronger flow due to the change in the flow splits.

Figure 14 contains an example plot of the Metal Temperature Factor (MTF) as a function of location. In general, the MTF data show a peak value at about the 200 mm location. However, the variation in location of the peak MTF does range from about 120 mm to about 250 mm depending on the flame structure which is being affected by the hardware configuration, air flow rate, and nitrogen flow rate. The location of the peak MTF probably represents a stagnation point associated with the major recirculation zone.

III. RECOMMENDATIONS:

Suggestions for Follow-on Research.

There have been considerable insights into the operational characteristics gained from this summer's faculty and graduate student research program. There is much that yet needs to be done before a full understanding of the burner is achieved. There are several investigations that might be suggested to further these insights. Additional work needs to be done on the flow partitioning in the low and high swirl nozzles. The flow and pressure drop ranges need to be carried to higher values. Additional lean blow out measurements also need to be done, particularly at higher flow ranges. Work needs to be done to quantify the volume of a flame that is attached to the center of the burner and the amount of air and fuel that is actually entering the flame zone. This data is necessary in order to better determine the actual fuel equivalence ratio in the "well mixed" flame zone. Laser sheet lighted images (either MIE scattering from a

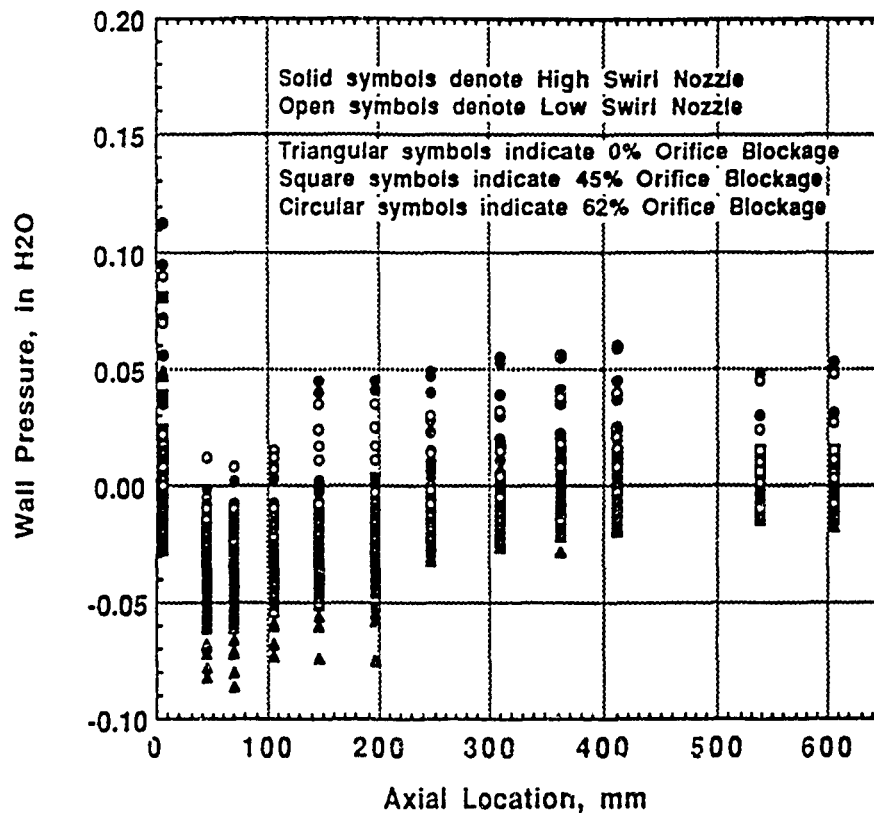


Figure 13 - Summary Graph Showing All Wall Pressure Data Taken.

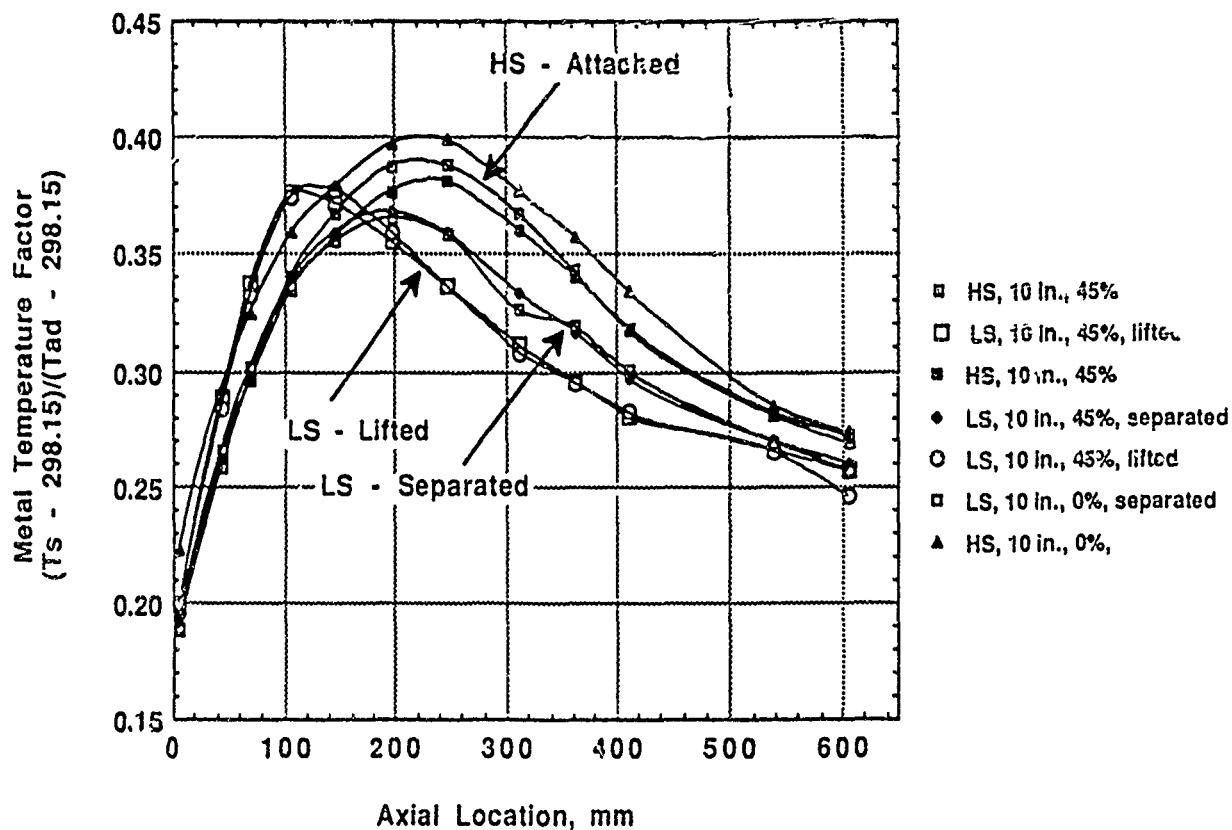


Figure 14 - Example of Metal Temperature Factors at 368 slpm (70 F) Air Flow

seeded flow, or LIF images of OH or NO) would be very useful to provide mappings of the flow fields. This information is essential in order to better understand where the flame fronts are located, and the characteristics of the eddies and other turbulent structures. Local gas temperature data and mixture fraction measurements would be very useful information for comparison to model predictions and similar data from either the Task 100 or Task 200 Combustors. The temperature data could be obtained using a CARS diagnostic system. Gas mixture fraction is a more difficult diagnostic, but there may be other diagnostics that could be used to measure the concentration of a seed gas (e.g. CO₂) such as degenerate four wave mixing. This area needs further investigation. The air flow rates to the primary and secondary swirlers were thought to have a major impact on the operation of the combustor. Much needed data on this characteristic could be obtained if the facility were modified so that the air flow rates to the various injector passages could be independently controlled.

Application for a Mini Research Initiation Grant.

An application for a mini research grant will be prepared as a result of the summer fellow research program. The proposed project would include the measurements on the Task 150 provided to BYU and installed under a research initiation grant during this past year. This burner is identical to the one tested in this summer at the Aero Propulsion and Power Laboratory, Wright Patterson AFB during this fellowship project. The burner incorporates the high swirl injector supplied by Pratt and Whitney Aircraft Co, East Hartford, Connecticut. The proposed study will include: additional measurements of lean blow out, additional flame characterization measurements using film photographs and video images, CARS gas temperature measurements, and sheet lighted film photographs from MIE scattering from a seeded flow. The Aero Propulsion and Power Laboratory, Wright Patterson AFB has companion programs, the Task 100 burner program, and the Task 200 burner program. The Task 100 burner was the burner investigated during the previous (1990) summer program. The Task 200 burner incorporates four swirling burners in a linear array. The proposed project would bridge the Task 100 and Task 200 programs and would provide much needed data that can not be obtained in any other burner. The proposed BYU project would fill a need to investigate the swirling burner in a single burner configuration. In addition, to basic combustion measurements, a future program at BYU could begin to investigate the formation of NO_x pollutants at the higher fuel equivalence ratios associated with higher performance gas turbines.

ACKNOWLEDGEMENTS

I wish to thank the Aero Propulsion and Power Laboratory, Wright Patterson Air Force Base, and the Air Force Office of Scientific Research for sponsorship of this research. The timely help of Mr. Gary Moore and other individuals at Research and Development Laboratories, Inc. is greatly appreciated. Special recognition of the help of Mr. Milton H. Danishik (RDL) is acknowledged. Not only did he provide

extraordinary help in providing information about living accommodations and base information, but he became a good friend as we shared the same office.

The very successful experience gained during this summer research program is in large measure due to the very excellent support of my sponsor, Dr. W. M. (Mel) Roquemore (WRDC/POSF). Special thanks are also given to Mr. Dale Shouse and Mr. Melvin Russell who provided much needed support with instrumentation and the burner and facility modifications. The very well qualified research staff at Systems Research Laboratory, Inc. (SRL) is greatly appreciated. Special thanks are in order for Dr. Larry Goss, Mr. Benjamin Sarka, and Mr. Michael Post of SRL. Their willingness to help, to answer questions, and the help that they provided made this summer's effort truly rewarding.

IV. REFERENCES

Lewis, B., and G. von Elbe, (1987), Combustion, Flames and Explosions of Gases, 3rd Edition, Academic Press, Inc.

Longwell, J.P., E.E. Frost, and M.A. Weiss, (August 1953), "Flame Stability in Bluff Body Recirculation Zones", Industrial and Engineering Chemistry, 45(8), pp 1629-1633

Longwell, J.P., and M.A. Weiss, (August 1955), "High Temperature Reaction Rates in Hydrocarbon Combustion", Industrial and Engineering Chemistry, 47(8), pp 1634-1643

Sturgess, G.J., D.G. Sloan, A.L. Lesmerises, S.P. Henneghan and D.R. Ballal. (June 1990), "Design and Development of a Research Combustor for Lean Blowout Studies", 35th International Gas Turbine and Aeroengine Congress and Exposition. Brussels, Belgium.

MARK WILLIAMS

REPORT NOT AVAILABLE
AT TIME OF
PUBLICATION

FINAL REPORT
TWO DIMENSIONAL SIMULATION OF RAILGUN PLASMA ARCS
Submitted to
1991 US AFOSR - RDL SUMMER FACULTY/
GRADUATE STUDENT RESEARCH PROGRAM

Research Period: June 20, 1991 to August 28, 1991

Faculty Research Associate: Dr. Manuel A. Huerta
Professor of Physics
University of Miami
Physics Department
Coral Gables, FL. 33124

Sec. Sec. No. 262-70-9459
Telephone No. (305) 284-2323, Ext.8

Graduate Student Research Associate:
Mr. G. Christopher Boynton
University of Miami
Physics Department
Coral Gables, FL. 33124
Telephone No. (305) 284-2326, Ext.6

Research Location: WL/MNSH
Eglin AFB, FL 32542-5000

Effort Focal Point: Mr. Kenneth K. Cobb
WL/MNSH

Signed: Kenneth K. Cobb

Signed: Manuel A. Huerta
Manuel A. Huerta
Faculty Research Associate

Signed: G. C. Boynton
G. Christopher Boynton
Graduate Student Research Associate

Date: August 27, 1991

TWO DIMENSIONAL SIMULATION OF RAILGUN PLASMA ARMATURES

by

Dr. Manuel A. Huerta, Professor of Physics and Mr. George C. Boynton

ABSTRACT

Our code uses the equations of two-dimensional resistive MHD with Ohmic heating and radiation heat transport to simulate the internal dynamics of a railgun plasma armature. All quantities are advanced in time using an explicit Flux Corrected Transport scheme. We have done some theory to describe the initial fuse explosion and have modified the simulation to have a more realistic initial state. We now allow the driving current to be input from a data file. We also have developed a version that computes only one half of the rail to rail distance and forces mirror symmetry for the other half to save running time. We have done the background work to use an electrical conductivity that accounts for nonideal effects and to include turbulent viscous drag. We also studied the problem of the lubrication and drag of the projectile against the rail walls and found that it could be made to have little effect. A good deal time was spent doing a review of a new scheme for pulse radiation.

I. INTRODUCTION

I have spent the summer of 1991 working at WL/MNSH site A15, together with my graduate student G. C. Boynton who was supported under the 1991 USAF Graduate Student Research Program. We have been working on a two dimensional simulation of armature plasmas. Our main objective this summer was to improve our model by incorporating a better initial state, allowing for a time varying current from a data file, including nonideal electrical conductivity, and field augmentation, and examining the problem of viscosity.

We enjoyed abundant access to the VAX 8650 at site A15. The Cray time that we were provided came a little late in the research period and was not fully utilized. An impressive railgun facility has been built on Okaloosa Island at site A15 and there is quite a bit of experimental activity on diagnostics of the arc plasma armature. We had very useful discussions with Mr. Kenneth Cobb and Dr. Eugene Clothiaux about their experiments. They provided us with the data files needed to include an experimental current profile in our simulation. We also profited from discussions with Dr. Glen Rolader of SAIC regarding many aspects of the problem.

II. WORK ACCOMPLISHED

In our previous simulations we took the total current to be a constant in time and the initial state of the plasma was closely related to the equilibrium calculated by Powell and Batteh¹. The new simulation has an initial state in which the plasma starts in a much more dense state that is related to the thin exploding foil that produces the initial plasma. We did a good bit of theoretical modeling of the different stages as the foil melts, vaporizes, and ionizes. This would be useful if we decide to pursue the matter of foil initiation more fully. As it is the details are not important right now because the simulation takes the initial plasma to occupy only one computational cell. This is due to running time limitations and leads to large gradients, as shown in the current profile of Fig. 1. FCT is good at handling this sort of thing, however, and the code runs fine. Figs. 2 and 3 show how the plasma expands nicely.

The time varying current profile that we have used is derived from the experiments of Clothiaux and Cobb on their capacitor driven railgun. The profile is shown

in Fig. 4. The current drops as a capacitor module discharges and rises again as another module cuts in. As expected the plasma grows toward the rear when the current begins dips. The reason for this is that the current drop is felt mainly at the rear of the plasma due to the finite time required for magnetic field diffusion. The rear region begins to fall behind the main arc because a lower current means a lower magnetic force on it, so its acceleration becomes less than that of the main arc.

The rapid growth of the arc toward the rear has caused a problem in our simulation, but we can easily fix it. Our initial conducting arc is only one cell long but our computational region is much longer and mostly nonconducting initially. We do this so we can monitor the growth of the conducting region of the arc. The problem is that the way our code runs now the information of the true value of the total current is lost if the entire computational region becomes conducting. In effect we not only lose current out the rear of the computational region, but we lose contact with the true value of the total current and of the rear magnetic field. A good fix of this problem will be to change our boundary condition at the rear of the computational region so the magnetic field there is made to always follow the true magnetic field at the breach. This would tend to introduce a magnetic field gradient (read current density) right at the rear of the computational region, but it would diffuse rapidly because this region is not a good conductor. This kind of trouble highlights the importance of any drop in the total current, as expected.

Due to the rear boundary condition problem the acceleration starts to drop off as shown in Fig. 5. This drop off is not valid. One of our motivating factors in introducing the time varying current was to see if we could reproduce the buffer zone results of Cobb and Clothiaux, so far we have not been able to.

Our simulations so far have used the well known Spitzer² conductivity

$$\sigma = \frac{2}{Z\sqrt{m_e}} \left(\frac{4\pi\epsilon_0}{e} \right)^2 \left(\frac{2k_B T}{\pi} \right)^{3/2} \sigma_{SP}^* \quad \text{with} \quad \sigma_{SP}^* = \frac{\gamma_E}{\ln \Lambda}$$

where γ_E is of order unity and

$$\Lambda = \frac{12\pi}{Z_f Z \sqrt{n_e}} \left(\frac{\epsilon_0 k_B T}{e^2} \right)^{3/2}$$

is the Coulomb logarithm. We have decided to use the nonideal conductivity where σ_{SP}^* is replaced by σ_{KV}^* as discussed by Rolader and Batteh³. This can reduce the

conductivity by half at low temperatures which could have a significant effect on the current distribution in the plasma.

The problem of viscous drag was given a good deal of attention. We only have a small number (10-20) of cells in the rail to rail direction due to the need to reduce the running time. It is hopeless to have enough resolution to truly do the viscous problem and have the boundary layer come out correctly. It becomes necessary to do some sort of matching of the boundary layer flow to the exterior flow. We use the Spitzer² viscosity for the plasma

$$\mu = 2.21 \times 10^{-15} \frac{T^{5/2} A_i^{1/2}}{Z^4 \ln \Lambda} \frac{\text{gm}}{\text{cm} \cdot \text{sec}},$$

where A_i is the ion atomic weight. For a typical velocity of 1 km/sec and length of 10 cm the Reynolds number

$$R = \frac{\rho U l}{\mu} \approx 2.5 \times 10^8$$

which is a very large value. This puts us in the regime where there is a turbulent boundary layer. For this reason we use the empirical equations for the shear stress due to viscous drag in turbulent pipe flow taken from Schlichting⁴

$$\tau_0 = \frac{1}{2} \rho U_\infty^2 c'_f \quad \text{where} \quad c'_f = 0.0592 \left(\frac{U_\infty x}{\nu} \right)^{-1/5}.$$

We integrate this stress along the rails and subtract the total drag force from the forward force that acts on the projectile.

We also considered the problem of the viscous flow that lubricates the motion of the projectile along the rail. This is a generalization of lubrication flow⁴ because here the pressure is different behind and in front of the projectile. The lubrication problem is coupled to the projectile deformation due to the stresses caused by the force that gives the projectile its acceleration a_P . Say we have rigid rails and a projectile of square cross-section w^2 , length H and density ρ_P . Let the gap between the rail and the projectile have a width h_1 at the front of the projectile and h_2 at its rear. We simplify by letting the pressure in the lubrication flow have an average value P_L . The quantities

$$h_1 - h_2 \approx \frac{\nu_P \rho_P g H}{E - 2\nu_P G} \frac{w}{2} \quad \text{and} \quad h_1 \approx \frac{w}{2} \frac{P_L}{E}, \quad (1)_{h_1 h_2}$$

where E is the Young's modulus, G the shear modulus and ν_P the Poisson ratio of the projectile, are crucial in determining the flow in the gap. Let μ be the viscosity of the fluid in the gap and U be the speed of the projectile. The average pressure in the gap is

$$p_L \approx \frac{6\mu U H}{(h_1 - h_2)^2} \left(\ln\left(\frac{h_1}{h_2}\right) - 2\frac{h_1 - h_2}{h_1 + h_2} \right)$$

From this and Eq. (1) we are able to get h_1 and h_2 , which come out quite small, of the order of micrometers. The drag force due to this flow turns out to be quite small. The volume rate of fluid flow toward the rear of the projectile is

$$Q = \frac{wh_1h_2}{h_1 + h_2} \left(U - \Delta p \frac{h_1h_2}{6\mu H} \right)$$

where Δp is the pressure difference between the back and the front of the projectile. A negative Q would indicate a plasma blow by. Typical values of Q can be made small enough to be unimportant.

Finally a good deal of time was spent preparing a review of a proposal for a new type of electromagnetic pulse generator called an impulse gun.

IV. CONCLUSIONS

Our code has shown the potential to describe the main experimental facts and is now approaching a stage where it can be compared with experiments. This should give valuable insights into the workings of plasma armatures.

V. ACKNOWLEDGEMENTS

We would like to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsorship of this research. We would also like to thank Research & Development Laboratories for the efficient way in which they administered the USAF - RDL Summer Faculty and Graduate Student Research Programs.

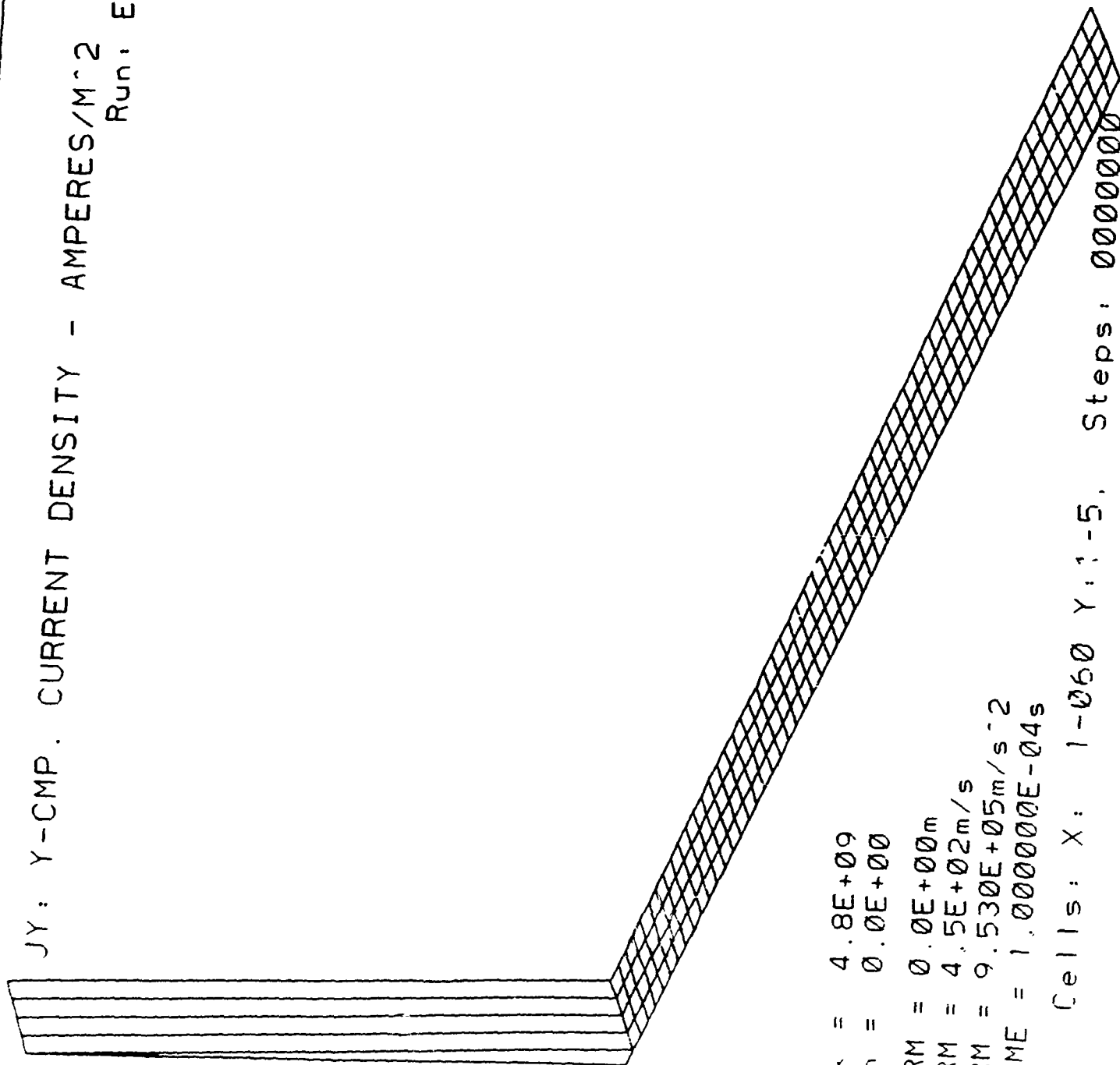
Our experience was productive and enjoyable thanks to the cooperation we received from many people at WL/MNSH. We would specially like to thank Mr. Kenneth K. Cobb who helped us so much in getting access to the computer resources that we needed. We also thank him for useful discussions of work in rail guns. We had fruitful discussions with several people, specially Dr. Glen Rolader of SAIC,

Dr. Eugene Clothiaux, and Dr. Chia Thio. We thank Mr. Ed Taconis for his support and Mr. Andy Marino for setting us up with abundant disk space and other resources on the VAX 8650 at site A15.

REFERENCES

1. Powell, J. D., and Batteh, J. H., "Plasma Dynamics of an Arc-Driven Electromagnetic Projectile Accelerator," J. Appl. Phys., vol. 52, 2717 (1981)
2. Spitzer, L., "Physics of Fully Ionized Gases", Interscience Publishers, 1962.
3. Rolader, G. E., and Batteh, J. H., "Thermodynamic and Electrical Properties of Railgun Plasma Armatures", IEEE Trans. Plasma Sci., vol. 17, pp. 439-445, June 1989.
4. Schlichting, H., "Boundary Layer Theory", McGraw-Hill, 1960, Chapter XXI.

JY: Y-CMP. CURRENT DENSITY - AMPERES/M²
 Run: Elmo

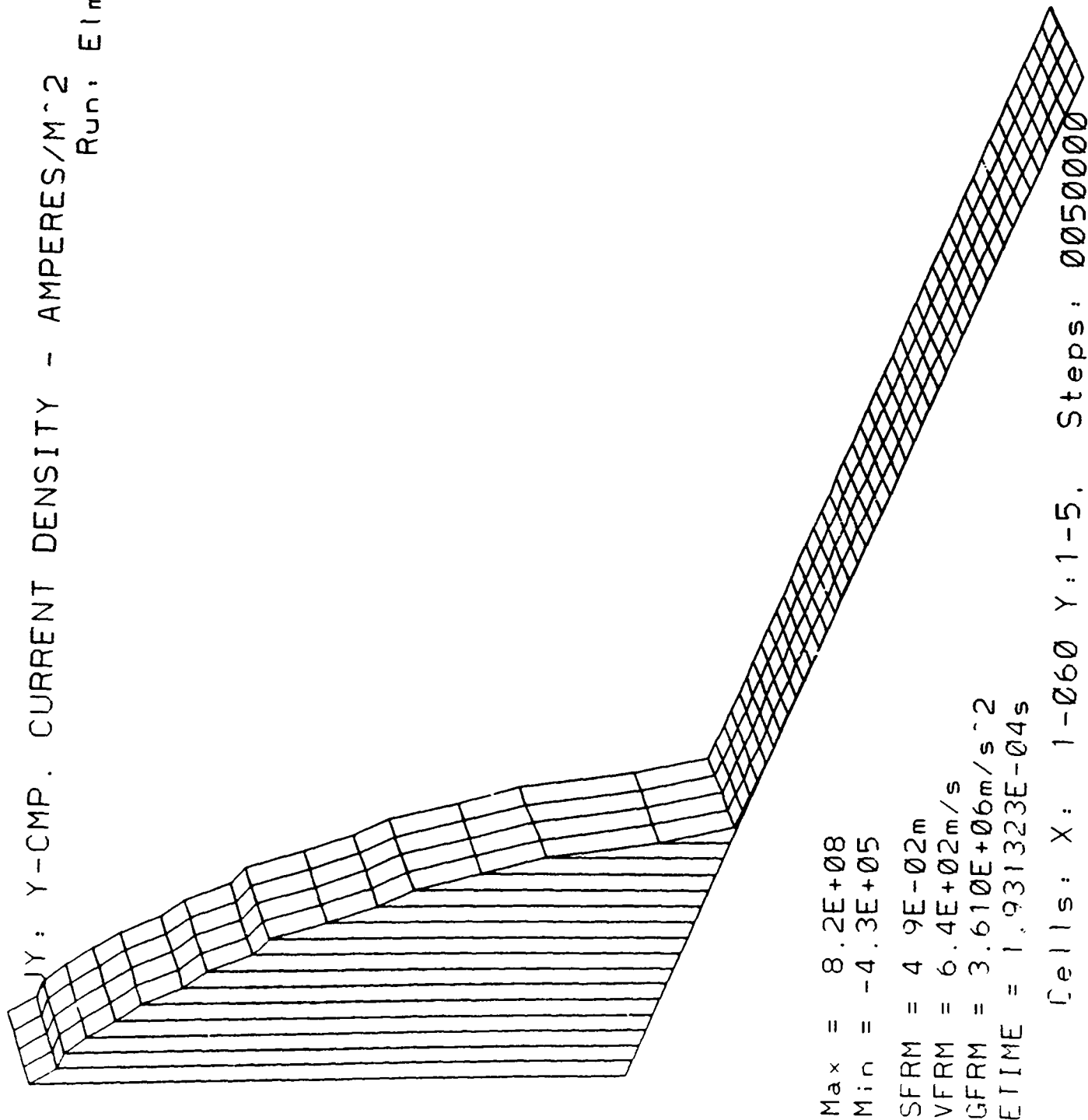


Max = 4.8E+09
 Min = 0.0E+00
 SFRM = 0.0E+00m
 VFRM = 4.5E+02m/s
 GFRM = 9.530E+05m/s²
 ETIME = 1.000000E-04s

Cells: X: 1-060 Y: 1-5, Steps: 000000

FIGURE 1

JY: Y-CMP. CURRENT DENSITY - AMPERES/M²
 Run: Elmo



Max = 8.2E+08
 Min = -4.3E+05
 SFRM = 4.9E-02m
 VFRM = 6.4E+02m/s
 GFRM = 3.610E+06m/s²
 ETIME = 1.931323E-04s

Cells: X: 1-060 Y:1-5. Steps: 0050000

FIGURE 2

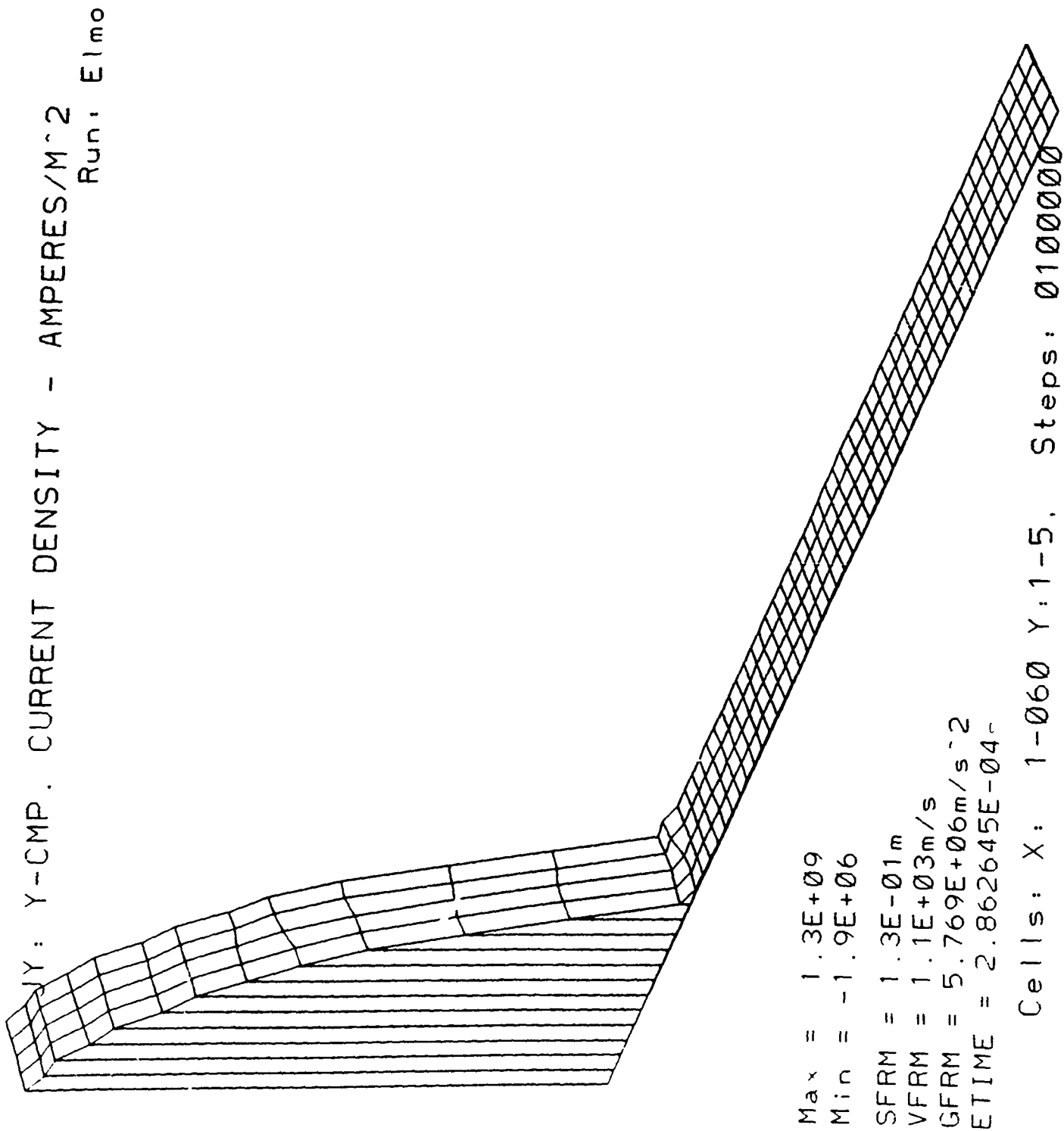


FIGURE 3

FIGURE 4

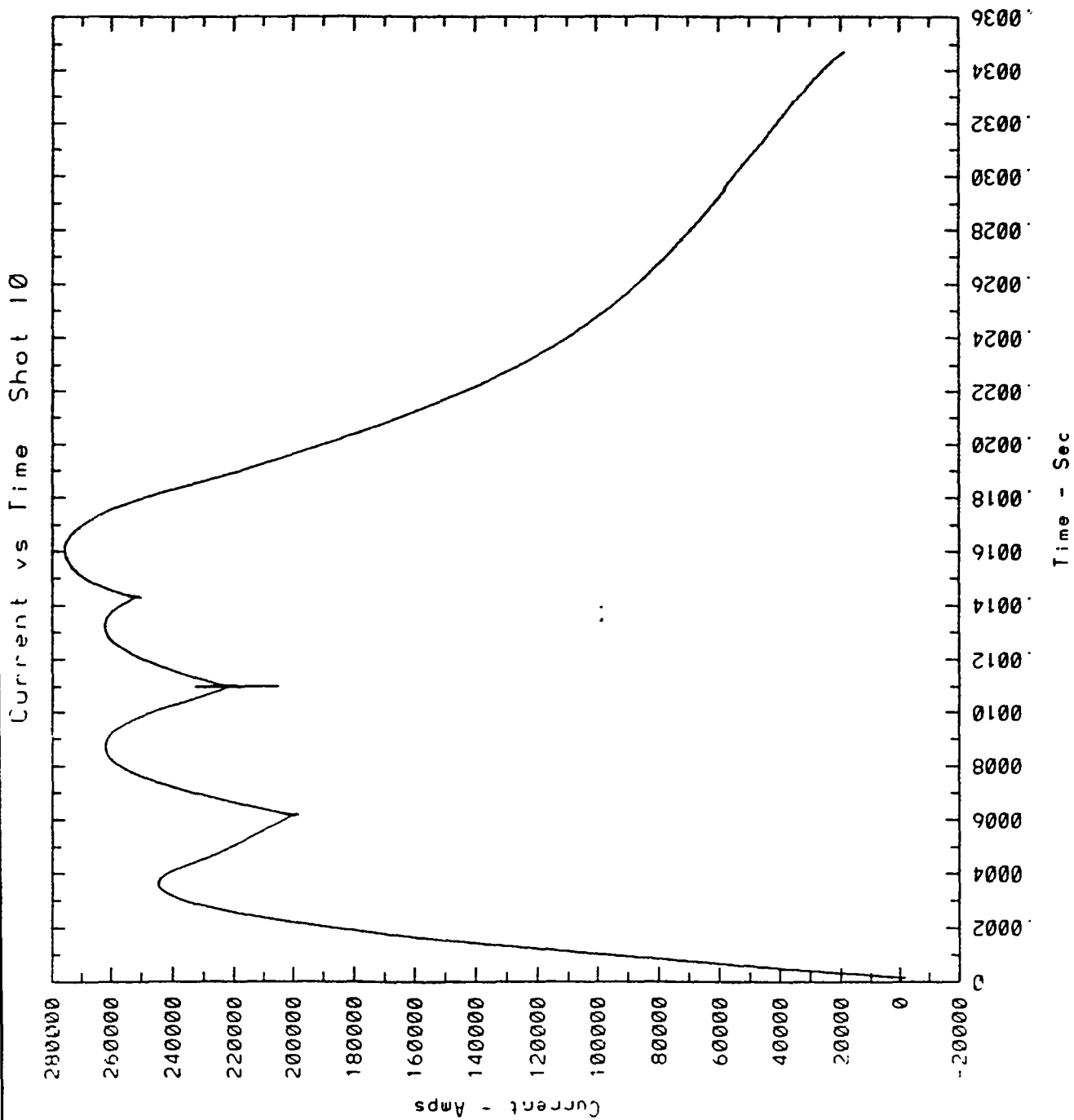
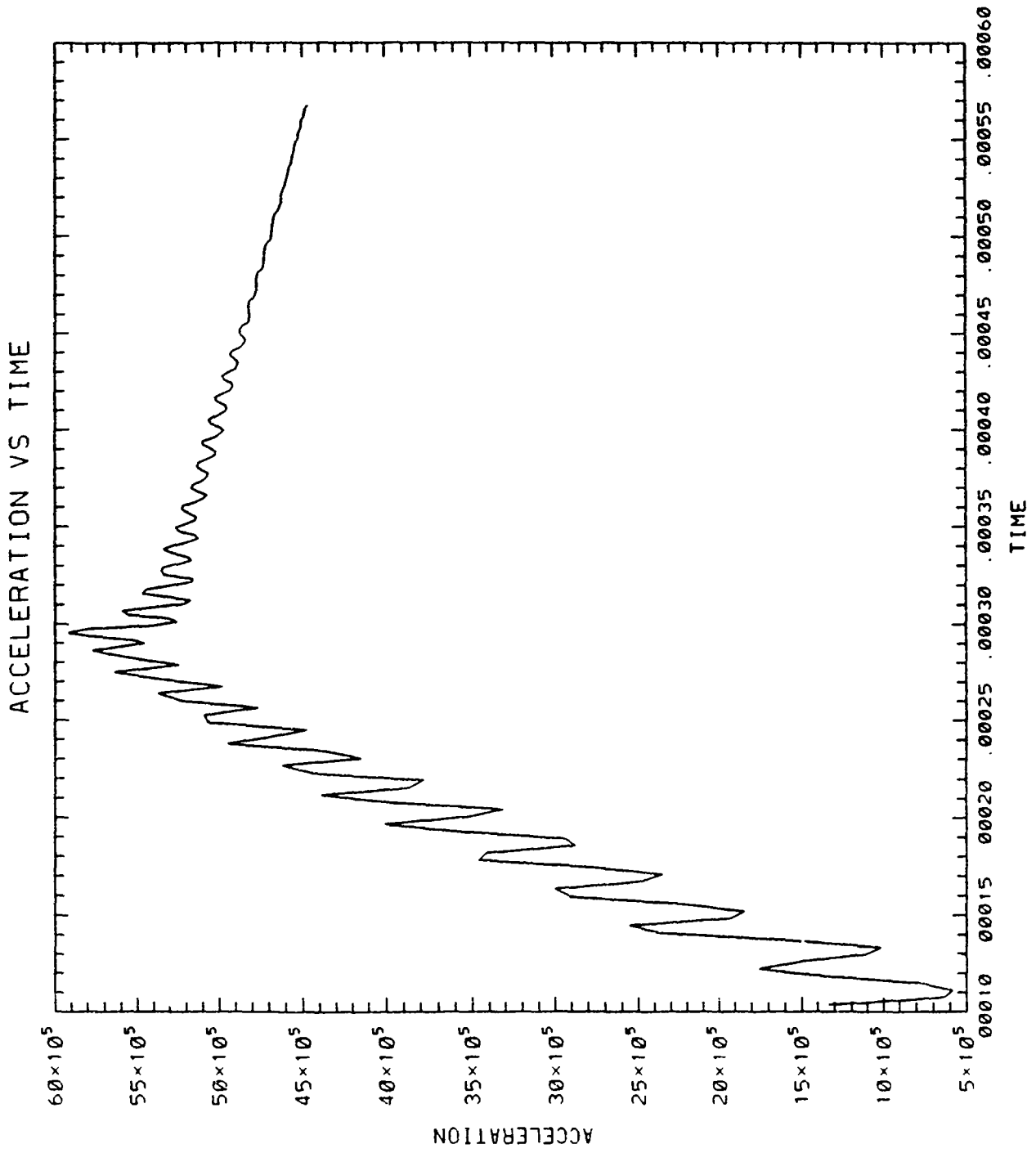


FIGURE 5



SETUP OF MULTISPECTRAL RADIOMETER SYSTEM

Laurence Flath

ABSTRACT

Recent studies¹ have found that state-of-the-art electronic video technology is superior to film-based high-speed imaging of in-flight events (missile releases, reconnaissance, etc.) in many respects, including greater amounts of useful data and shorter time from capture to data analysis. As a first step in constructing a high-resolution, high-speed video system, quantities like the nominal scene radiance of such events need to be found. A multispectral radiometer was constructed over the past year for just this purpose at the University of Arizona's Optical Sciences Center, with funding by USAF WL/MNGI. With the basic radiometer completed and returned to Eglin AFB, the setup of data collection equipment and calibration of the system remained and has thus constituted the main focus of this summer project. In addition, analysis of current video-camera technologies was also carried out as part of further research on the high-speed video project.

INTRODUCTION

Present high-speed imaging of in-flight events (missile releases, reconnaissance, etc.) is a complex process². Film cameras operating at 400 fps (frames per second) are mounted under the wing of a plane with up to 1000 ft. of 16 mm film. Settings like the camera orientation, aperture stop, and shutter speed are preset on the ground; the pilot is only given control over the shutter release. Once initiated, the camera mechanisms take several seconds to get up to speed, after which the entire spool of film is exposed. At this point, there is no way to know whether the exposure was proper, i.e. if the camera saw the event, if there was too

much or too little light, etc. Once on the ground, the film is developed, a copy is made, and the copy is sent to a digitizer that transfers the film images into a digital format that computers can read and analyze. Although film (by itself) can record more information (represented by the MTF or Modulation Transfer Function) than current video cameras, the combination of unknown exposure, duplication, and digitization make state-of-the-art high-resolution video cameras' better than the complete film-based system. In addition, a high-speed, high-definition video camera system has the advantage of real-time (or near real-time) display of image data quality; the pilot and/or ground controllers can see what the camera is looking at, quite unlike the film system which can take anywhere from days to months for data display. The advantage of reduced turn-around time with the same or better data quality is alone enough to justify the development of a high-speed, high-resolution video system.

In order to develop specifications for such a system, the properties of typical in-flight scenes need to be known. Of special consideration are measurements of radiance and spectral content of such events. In the past year, WL/MNGI (D. Snyder) contracted the Optical Detection Labs at the Optical Sciences Center, University of Arizona (Dr. E. Dereniak) to build a flight-ready multispectral radiometer for just this purpose. The instrument was completed and shipped to Eglin AFB, where the final calibration and interfacing with a computer were to be carried out as this summer project. Using an AT-class PC and software, data collected by the radiometer are displayed and easily saved for later analysis.

MULTISPECTRAL RADIOMETER

The multispectral radiometer is a self-contained unit which is made to fit inside a standard instrument pod mounted under the wing of an airplane. Figure 1 shows the shape of the aluminum unit, which is roughly 24" x 12" x 8" and weighs

approximately 80 lbs. Although named a radiometer, the device really consists of four separate instruments, including two radiometers and two CCD cameras.

The arrangement (shown in Figure 2) of a Sony CCD camera, a 4° NFOV (Narrow Field Of View) spot radiometer, and a 24° WFOV (Wide Field Of View) radiometer is due to the setup of the Xybion MSC-02 multispectral camera system. A filter wheel that rotates in front of the CCD camera contains six interference filters spectrally spaced at 400 nm, 450 nm, 500 nm, 600 nm, 700 nm, and 800 nm. The Xybion camera was modified to include two photodiode radiometers so that what the CCD camera saw through a particular filter (wavelength) could be measured quantitatively; the WFOV radiometer essentially measures the radiance of the camera's entire field of view, whereas the NFOV radiometer measures only a small portion of that field. This summer, a compact Pulnix CCD camera was mounted on the NFOV's lens bracket so that the scene could also be imaged without filters obstructing the view.

Custom electronics were fabricated so that data from the radiometers could be overlaid on the video signal generated by the Xybion camera. A block diagram of the radiometer is shown in Figure 3. Video and data cables link the unit to a Datum Video Data Encoder (VDE), which overlays the digital radiance data, as well as IRIG B time, onto the CCD video signal. Figure 4 shows a typical output of the VDE. In practice, the multispectral radiometer, the VDE, and a VCR to record its output are all mounted in the aircraft. Video tapes can hold hours of imagery and its corresponding radiance data in a compact package, with the convenience of random-access data retrieval and analysis.

INTERFACING THE MULTISPECTRAL RADIOMETER

With the majority of hardware completed, providing a means for easy display and storage of data and statistics was the next task. Special hardware could have been built for

this purpose, but would lack the simplicity and flexibility offered by an interfaced personal computer.

The means of data collection decided, an input/output (I/O) card for the PC was acquired: Data Translation's DT2817, a programmable 4-I/O port interface card. Each 8-bit port is programmable as either an input or an output. A parallel cable including the eight data bits, sync and handshaking lines was run from the Datum Video Data Decoder to the I/O card.

At first, the software used to control the data input from the video data decoder was meant to be small and utilitarian. As such, handshaking routines were written in Microsoft's QuickBASIC. The anticipated problem with this approach was that $6 \text{ bytes/line} \times 10 \text{ lines} = 60 \text{ bytes}$ of data needed to be clocked out of the decoder every 16.7 ms (one video field). It turned out that the compiled QB code was just fast enough to accomplish this, with 3 ms to spare. Of course, this left only 3 ms left to do real-time calculations on the data and display it on the screen (not feasible). Even so, this was not the reason for going to another programming language; the real problem arose when a large number of frames' data needed to be stored real-time in the computer's memory. QB did not have memory models that supported data arrays with the large indices needed to calculate the radiometer radiances quickly. Sophisticated memory management like segment switching could have been used, but the overhead required for such techniques would have exceeded the minimal amount of time available to do anything useful.

It was very fortuitous that a mini-course on C was offered at Eglin, for C was just the language needed to circumvent the problems with BASIC. The first stage was to implement what had been done in QuickBASIC in Borland's TurboC. The resulting code executed much faster, leaving over 10 ms of free time per field to store and make calculations on the incoming data. The huge memory model in

TurboC allowed for 600 fields' worth of digital data, timecodes, and radiance values; at 60 fields/sec, this equates to 10 seconds of consecutive data being stored in the computer's memory.

Once the feasibility of data acquisition was determined, much time and effort was spent on creating an easy-to-use, aesthetically pleasing interface. The reasoning behind this was that unfriendly software tends not to be used; my intention was to make users want to gather data with the radiometer. Color, menus, real-time bar-graphs of numerical data, and a consistent interface were just a few of the methods used in accomplishing this goal.

The core function of the program DATACOL.EXE, is the collection of multispectral data from the Datum video data decoder. Two methods of collection are available to the user; real-time display of data and recorded data. The real-time option allows for continuously changing values to be displayed onscreen for all six filters' radiances. The corresponding frame's IRIG B time, 16-bit digital radiometer data and calculated radiance for both the NFOV & WFOV are updated about twice per second onscreen. Conversely, the recording mode samples and stores to memory up to ten seconds' worth of consecutive data, and then allows the user to scroll through the data using the cursor keys. The radiances are calculated from a sixth-order polynomial fit to calibration data previously taken on the radiometer. The coefficients for this polynomial are stored in a file that can easily be modified if another calibration is done on the unit.

Another reason for the software programming taking a long time to complete was the addition of many features not originally slated for this "simple" program. Thanks to the code being written in a fairly compartmentalized fashion, making such additions was relatively simple. Some diagnostic utilities such as viewing real-time and recorded raw data (bytes) from the video data decoder were included, as well

as a statistics screen showing the averages and standard deviations of the recorded multispectral data. A module to store all of the recorded times, data, radiances, and statistics on disk was also added.

Attached at the end of this report is a copy of the software manual that remained with the multispectral radiometer data collection system at Eglin AFB. It describes the basic operation of DATACOL.EXE, and shows what typical screens look like during execution of the program.

FLIGHTLINE TEST OF MULTISPECTRAL RADIOMETER

On August 7, 1991, Donald Snyder (WL/MNGI), Howard McCormick (WL/MNGI), Jeff Rowe (Sverdrup), and I conducted a test of the multispectral radiometer by bringing the complete system out to the Eglin AFB flightline. The procedures required to gain access to the flightline were quite involved due to security, but thanks to an aircraft which was grounded for routine maintenance, we were able to collect data on surfaces similar to those intended in the radiometer's initial design.

The radiometer and essential equipment was stacked in a cart that was easily moved around the aircraft. Power and video cables 100 feet long were brought to the cart from an air-conditioned truck, which housed a generator and the computer/video-decoder so that we could "see" the spectral data using the software written.

The radiometer was pointed at different surfaces on the plane over various lighting conditions (sun, shade, overcast, etc.). Over one hour of video was recorded simultaneously on two VCR's, one saving the data-encoded multispectral (Xybion) output, while the other recorded a normal video scene using the Pulnix camera. At sixty fields per second, this represents a huge amount of data to be analyzed, thus the results of this test will take time to compile. A full description of the flightline test will be the subject of a paper later this year.

CURRENT VIDEO TECHNOLOGIES

As mentioned previously, the multispectral radiometer is only a small part of the high-speed video development project of WL/MNGI. Research is currently being conducted in many areas, including focal plane arrays (CCD), A/D conversion, gigabyte memory caches, digital data compression, storage, and display. As an introduction to the facilities and equipment available to me, analysis of the Pulnix 745E CCD camera was conducted with Dr. Eugene Chenette (University of Florida, EE). Looking at the raw video coming out of the Sony CCD head, we were amazed at the quality of the final output signal. It turns out that a Sony image-processing chip samples and amplifies the weak, noisy, and heavily dc-biased output of the CCD in a truly admirable manner. With only a crude block diagram of the chip from Sony, we were able to determine that the chip uses a variation of double-correlated sampling to remove noise and amplify the minute changes in pixel-output that represent orders of magnitude change in incident photon-flux. If such a method (or even a commercially available, domestic chip) could be adapted to the focal plane arrays we are analyzing for the high-speed, high-resolution camera system, the need for subsequent data processing circuitry might be eliminated. In general, image-processing on the analog video signal requires far less hardware and is much faster than comparable digital signal post-processing, but has the disadvantage of adding noise to the signal.

Many other performance studies on high-speed CCDs, A/D converters, and signal-processing methods were conducted under the direction of Dr. Chenette. With this extra experience and knowledge, more informed decisions as to the form and construction of a high-speed, high-resolution video testbed can be made.

CONCLUSION

The need for digital high-speed, high-resolution video systems has been proven, and is growing steadily as computers become more involved in image processing. For working systems to be constructed, a testbed for electronic camera components is crucial. The first part of such a system is a multispectral radiometer to measure spectral light levels, from which the required CCD dynamic ranges can be determined. A radiometer built at the University of Arizona's Optical Sciences Center was tested and performed very well in combination with a PC and software interfaced to it. The multispectral radiometer will be a useful instrument for many years to come, as extensive data can be analyzed with a simplicity and utility not previously available.

ACKNOWLEDGEMENTS

Many people aided my summer project at Eglin AFB who need to be given credit for all of their help and support: Donald Snyder (WL/MNGI), Dr. Eugene Chenette (UF), Howard McCormick and his technicians (WL/MNGI Optics Lab), Jeff Rowe (Sverdrup), and Martin Sensiper (UCF). At the Optical Detection Labs at OpSci (U of AZ), my advisor Dr. Eustace Dereniak, Thorsten Graeve, Davis Lange, and Michael Descour also require my thanks, as it was their work which made the radiometer from conception to delivery.

REFERENCES

- 1,2 D. Lange, "High Speed Video Systems for Munitions Testing", AFOSR Final Report, 1990.
- 3N. Carlton, E. Chenette, J. Rowe, "Characterization of High-Speed Video Systems: tests and analyses", report from Sverdrup Technology/TEAS Group to WL/MNGI, 1991.

FIGURES

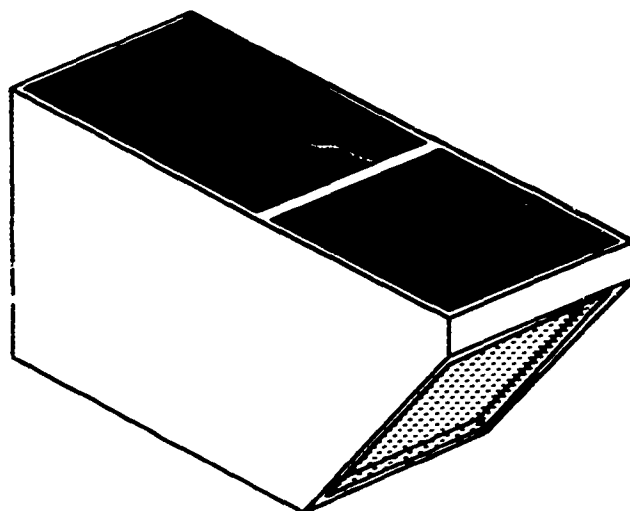


Figure 1. MultiSpectral Radiometer

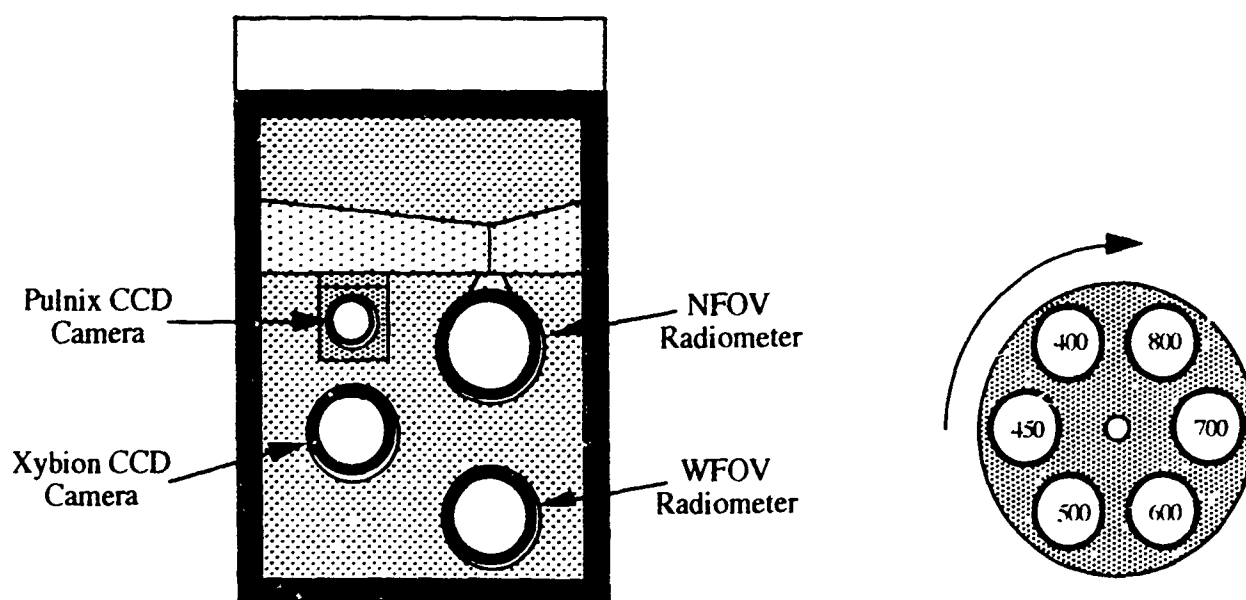


Figure 2. Arrangement of radiometer instruments

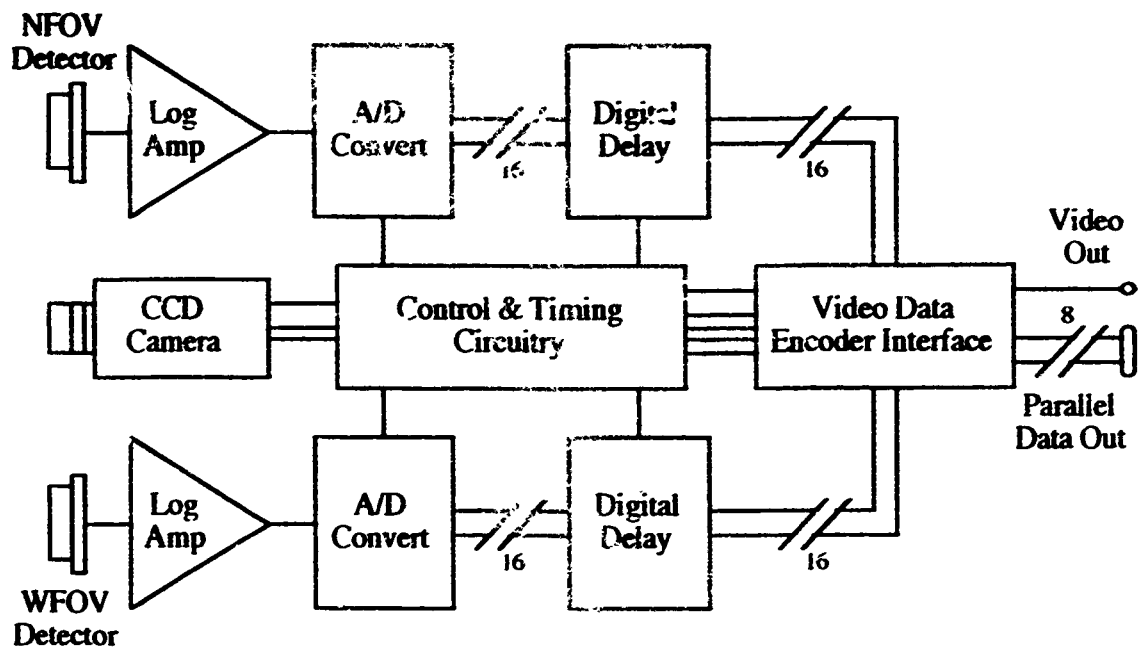


Figure 3. MultiSpectral Radiometer Internal Diagram

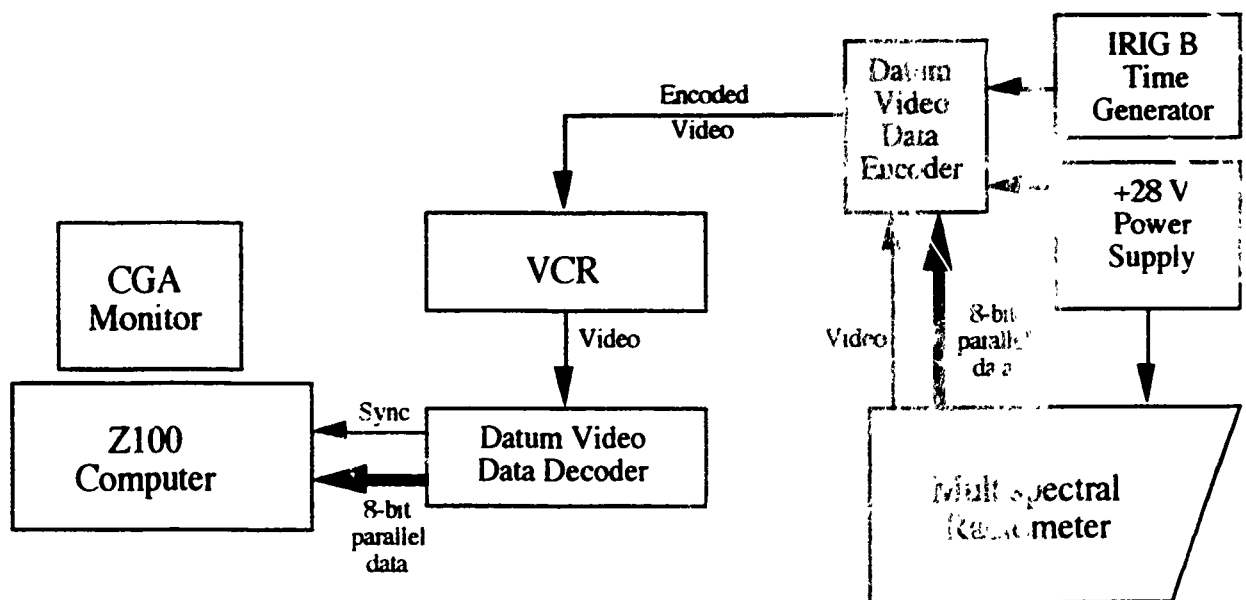


Figure 4. MultiSpectral Radiometer Data Collection System

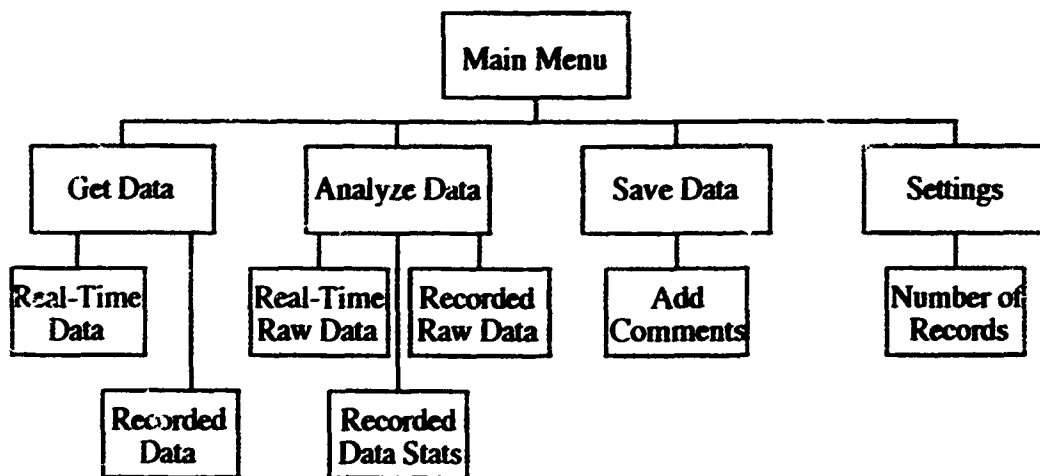


Figure 5. Diagram of DATACOL.EXE.

Manual for Multispectral Radiometer Data Collection Software DATACOL.EXE

System Requirements:

- Zenith Model Z100 '286 computer, with I/O card installed at Base Address = 0228H (See card manual).
- Datum Model 9200 Video Data Reader.
- Video containing 10 lines of 48-bit data; either live from multispectral radiometer or videotape of radiometer output.
- The following files in the \dc directory:

DATACOL.EXE Main data collection program.

RAD_COEF.DAT Coefficients of 6th-order polynomial relating digital number to radiance (See radiometer manual).

Running DATACOL.EXE:

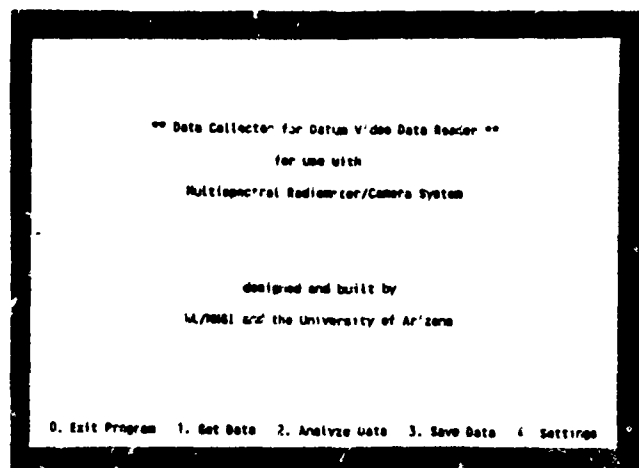
From root directory, type: dc

OR

From \dc, type: DATACOL

Main Menu Screen:

0. Exit program and return to DOS.
1. Enable data collection, both real-time and recording; display of decoded data on-screen.
2. Analyze both real-time and recorded data in hexadecimal mode; also display of statistics on recorded data.
3. Save recorded data to disk.
4. Change number of fields sampled when recording data.



Data Collection Screen:

0. Return to main menu screen.
1. Real-time display of data; automatically selected upon entry to data collection screen.
2. Record sequential fields of data (16.67 ms apart). After sampling, formatted data is displayed and the following cursor keys can be used to scroll through the data.

↑, PgUp, move up six fields (one filter-wheel rotation) i.e. backward $1/10$ sec in time.

Home, display first six fields.

↓, PgDn, move down six fields (one filter-wheel rotation); i.e. forward $1/10$ sec in time.

End, display last six fields.

** Reading Data Encoded in Fields **

Filter	Time	NFOV Data	NFOV Rad.	WFOV Data	WFOV Rad.
1 (400 nm)	0219:10:29:08.648	43300	2.485e+00	51044	3.868e+00
2 (450nm)	0219:10:29:08.765	48783	3.252e+00	33097	3.954e+00
3 (500nm)	0219:10:29:08.881	39958	4.206e+00	31899	4.37e+00
4 (600nm)	0219:10:29:08.998	39625	3.695e+00	31602	3.99e+00
5 (700nm)	0219:10:29:09.114	36133	2.816e+00	23559	5.349e+00
6 (800nm)	0219:10:29:09.231	32981	3.330e+00	18629	5.720e+00

Reading Fields: 4 Min NFOV Max Min WFOV Max

NOTES: - Data are digital numbers from radiometer; 0 is bright, 65535 is dark
- Radiances are in $\mu\text{W}/\text{cm}^2/\text{sr}$

0. Main Menu 1. Real-Time Data 2. Recorded Data

** Recorded Data from Fields **

Filter	Time	NFOV Data	NFOV Rad.	WFOV Data	WFOV Rad.
1 (400 nm)	0219:08:29:13.440	41923	2.895e+00	25572	5.958e+00
2 (450nm)	0219:08:29:13.477	40620	3.290e+00	26634	6.326e+00
3 (500nm)	0219:08:29:13.494	39524	4.352e+00	25588	7.070e+00
4 (600nm)	0219:08:29:13.510	38148	4.143e+00	25855	6.204e+00
5 (700nm)	0219:08:29:13.527	33010	2.977e+00	18083	8.134e+00
6 (800nm)	0219:08:29:13.544	29186	4.496e+00	13859	8.327e+00

Recorded Fields: 1 - 6 Min NFOV Max Min WFOV Max

NOTES: - Data are digital numbers from radiometer; 0 is bright, 65535 is dark
- Radiances are in $\mu\text{W}/\text{cm}^2/\text{sr}$

0. Main Menu 1. Real-Time Data 2. Record Data

Analyze Data Screen:

0. Return to main menu screen.
1. Hexadecimal display of real-time data; display corresponds to bytes encoded on live video.
2. Hexadecimal display of recorded data; display corresponds to bytes encoded on fields recorded to memory. The following cursor keys can be used to scroll through data:

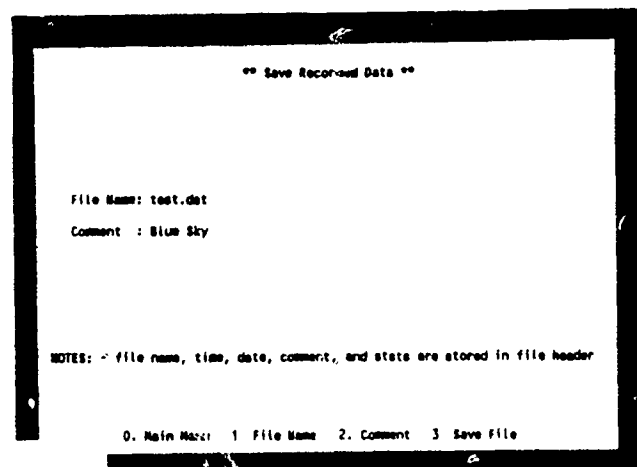
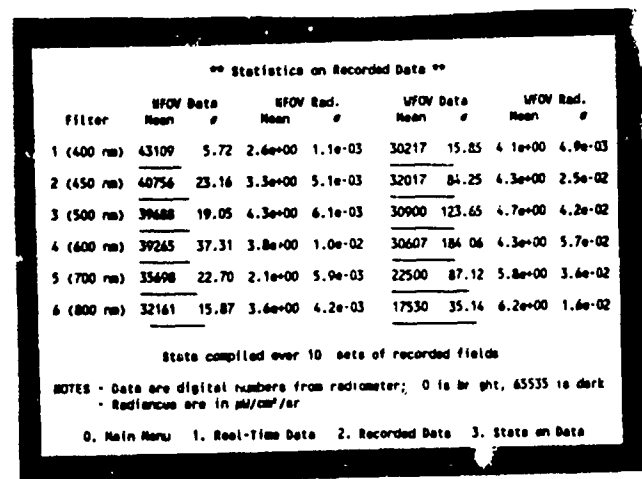
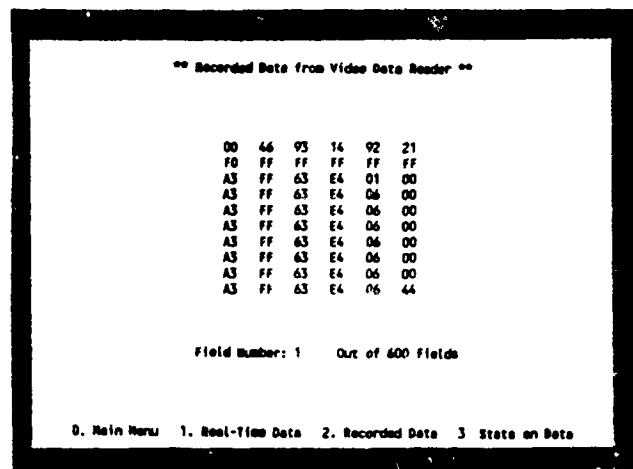
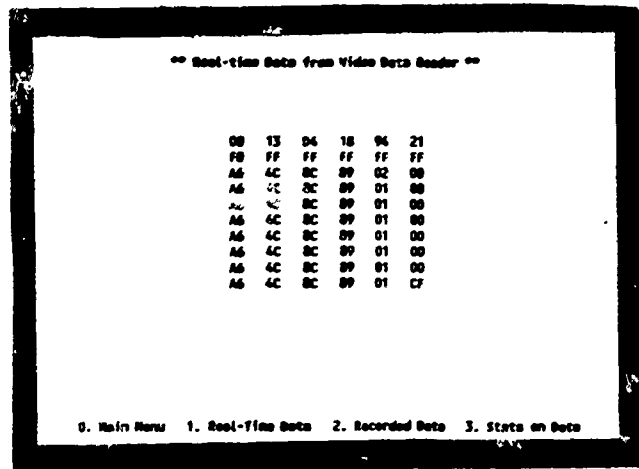
0. Main Menu 1. Real-Time Data 2. Recorded Data 3. State on Data

- ↑, move up one field ($1/60$ sec, previous filter).
- PgUp, move up six fields (one filter-wheel rotation): i.e. backward $1/10$ sec in time.
- Home, display first field.
- ↓, move down one field ($1/60$ sec, next filter).
- PgDn, move down six fields (one filter-wheel rotation); i.e. forward $1/10$ sec in time.
- End, display last field.

3. Display statistics on recorded data; computes average digital values and radiances, as well as standard deviations for both the NFOV and WFOV radiometers. In general, low σ 's in the data are better. The data σ 's represent the useful number of bits in the radiometers' outputs: e.g. $\sigma = 15.23$; $\log_2 15.23 \approx 4$; therefore, the number of useful bits = $16 - 4 = 12$ bits.

Save Data Screen:

0. Return to main menu screen.
1. Enter file name; mistakes can be corrected via Backspace and Del.
2. Enter comment to be inserted into data file, up to 80 characters long.
3. Saves file name, time, date, comment, data statistics, and recorded data to file specified in part 1.



Settings Screen:

0. Return to main menu screen.

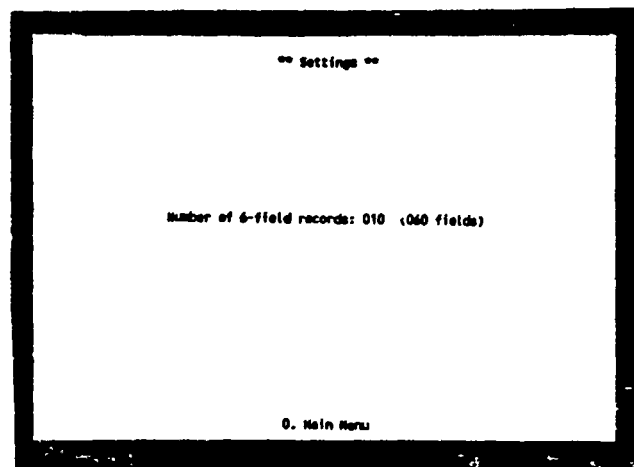
Use cursor keys to change the number of fields to sample when recording data:

↑, reduce the number of fields by one.

PgUp, reduce the number of fields by ten.

↓, increase the number of fields by one.

PgDn, increase the number of fields by ten.



Default setting is ten six-field records (filter-wheel rotations), which equates to sixty fields recorded (one sec.). Maximum setting is 100 six-field records (600 fields) or 10 sec. worth of recorded data.

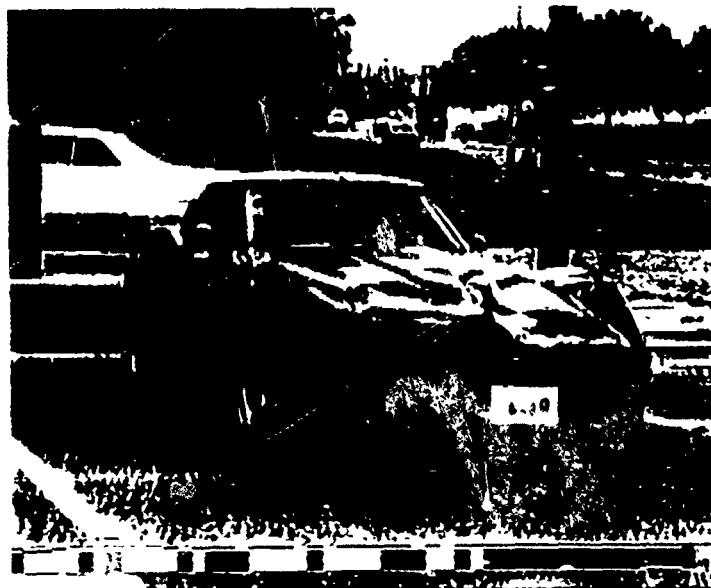
Formatting Notes:

Hex display of data corresponds to bytes encoded on video:

```

** Recorded Data from Video Data Reader **

IRIG B Time 00 46 93 14 92 21
             F0 FF FF FF FF FF
             A3 FF 63 E4 01 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 00
             A3 FF 63 E4 06 44
             NFOV WFOV Filter
             Number
Field Number: 1 Out of 600 Fields
  
```



RAD_COEF.DAT format: coefficients for 6th-order polynomial (specified in radiometer manual) are listed in a text file, one per line *exactly* 12 characters long; one sign, seven mantissa, "e", one exponent sign, two exponent. The first 6 lines of an example file:

```
+2.83987e+00    B0
+7.34684e-05    B1
-1.22382e-08    B2
+4.81734e-13    B3    400 nm, NFOV
-9.93230e-18    B4
+1.02016e-22    B5
-4.05779e-28    B6
```

This is followed by 450 nm, 500 nm, 600 nm, 700 nm, 800 nm seven-line sets of coefficients for the NFOV radiometer, and then a corresponding set of 42 lines of WFOV coefficients.

Data output file format is shown below in a typical file:

```
test.dat
Comment on scene.
Written at 0219:10:29:08
Filter Wdat_av Wdat_o Nrad_av Nrad_o Wdat_av Wdat_o Wrad_av Wrad_o
1 43305 7.47 2.6e+00 1.5e-03 31100 9.86 3.9e+00 2.8e-03
2 40782 13.08 3.3e+00 2.8e-03 33071 33.90 4.0e+00 9.1e-03
3 39956 11.02 4.2e+00 3.5e-03 31901 47.18 4.3e+00 1.5e-02
4 39640 15.83 3.7e+00 4.3e-03 31674 59.67 4.0e+00 1.7e-02
5 36121 20.12 2.0e+00 5.0e-03 23552 23.27 5.4e+00 9.0e-03
6 32958 19.50 3.3e+00 4.9e-03 18710 24.07 5.7e+00 1.0e-02
Filter Time NFOVdata NFOVrad WFOVdata WFOVrad
1 0219:10:29:08.748 43306 2.604e+00 31096 3.858e+00
2 0219:10:29:08.765 40783 3.252e+00 33098 3.954e+00
3 0219:10:29:08.781 39960 4.205e+00 31925 4.333e+00
4 0219:10:29:08.798 39660 3.685e+00 31729 3.959e+00
5 0219:10:29:08.815 36142 2.013e+00 23567 5.346e+00
6 0219:10:29:08.831 32996 3.326e+00 18714 5.682e+00
1 0219:10:29:08.848 43309 2.603e+00 31103 3.856e+00
2 0219:10:29:08.865 40792 3.250e+00 33069 3.963e+00
3 0219:10:29:08.881 39958 4.206e+00 31899 4.342e+00
4 0219:10:29:08.898 39644 3.690e+00 31675 3.976e+00
5 0219:10:29:08.915 36139 2.014e+00 23550 5.353e+00
6 0219:10:29:08.931 32969 3.333e+00 18721 5.679e+00
1 0219:10:29:08.948 43297 2.606e+00 31091 3.860e+00
2 0219:10:29:08.965 40756 3.258e+00 33025 3.975e+00
3 0219:10:29:08.982 39938 4.213e+00 31840 4.362e+00
4 0219:10:29:08.998 39625 3.695e+00 31602 3.998e+00
5 0219:10:29:09.015 36135 2.015e+00 23521 5.365e+00
6 0219:10:29:09.032 32971 3.332e+00 18681 5.697e+00
```

LASER POLARIMETRY

Graduate Student Researcher

Randy Gove

Abstract

Research in the field of Mueller matrix polarimetry was furthered with improvements in the data reduction and calibration technique, and construction of a visible light polarimeter. The data reduction and calibration technique was improved by implementing Mueller matrix equations which compensate for element orientational errors and nonideal retardances. These equations improved the accuracy of measurements made on the polarimeters. Infrared Mueller matrix measurements were made on a sample of cadmium telluride (CdTe) as a function of voltage applied across the crystal. The linear retardance for CdTe was found to increase linearly with voltage from 100 to 2400 volts. The electrooptic tensor coefficient was determined to be 6.10×10^{-12} . Measurements were made on a silicon wafer by rotating it between two polarizers. The linear retardance was determined to be 3.5 degrees and the linear diattenuation was determined to be less than 1%. A spectrum analyzer was used to find the power output for each of the infrared Edinburgh lasers tunable wavelengths. A Mueller matrix polarimeter operating at 632nm was constructed and calibrated.

Introduction

This paper is a summary of polarimetry work performed at Eglin Air Force Base (AFB) during the summer of 1991. Polarimetry is the branch of ellipsometry concerned with measuring the polarization state of a light beam and deducing the polarization properties of bulk materials. A polarimeter is an optical instrument used for the determination of the polarization of the light beam. Given the knowledge of how the polarimeter itself acts on the light, the polarization change produced by inserting a bulk sample into the polarimeter may be determined. The polarization properties a material has may be characterized with or without applied electric or magnetic fields.[1]

The Air Force Armament Laboratory at Eglin AFB has long been concerned with the development of electrooptic materials for the infrared. Characterizing the action of electrooptic materials on polarized light is an essential step in understanding how to use the materials on devices which modulate radiation. The process of characterization may also aid in choosing materials for specific applications and in designing new materials. Results of this process may point the way toward doping known materials to increase modulation efficiency or may indicate a particular crystal orientation or lattice spacing which produces better modulation. Improvements in infrared electrooptic modulators themselves will improve image processing systems, seekers, and other tactical devices.

Laser polarimetry has been a research project at Eglin AFB since the construction of the infrared laser polarimeter during the summer of 1988. The laser polarimeter was designed by Dr. Dennis Goldstein of the Advanced Guidance Division (MNG) at Eglin AFB. Mr. David Chenault constructed the polarimeter while working as an Air Force sponsored summer graduate researcher in 1988. A description of the operation of the polarimeter is contained in Mr. Chenaults final report to Universal Energy Systems

(UES).[2] Since its construction, calibration and data reduction work has been performed during the summers of 1989 and 1990 by Mr. Randall Hodgson and myself, respectively. A summary of this work is contained in our final reports to UES.[{3},{4}]

Discussion of the Problem

The first goal of this research project was to include Mueller matrix equations into the data reduction program that would allow for compensation of element orientational errors and nonideal retardances. This improved the accuracy of measured sample Mueller matrices. Infrared polarimeter measurements were made on a sample of CdTe with an applied electric field to determine its electrooptic tensor coefficient and polarization properties. The electrooptic tensor coefficient will give an indication of the effectiveness of a material as a modulator material. Measurements were also made on a silicon sample, without an applied electric field to determine if it exhibited any natural linear retardance.

Purchase of a new tunable Edinburgh infrared laser was made this past year. The output power was determined for each of the tunable wavelengths between $9\mu\text{m}$ and $11\mu\text{m}$. This information will allow for polarimeter measurements to be made at other known wavelengths. A visible polarimeter was constructed from ground up and calibrated. The visible polarimeter is of the same general design as the infrared polarimeter but uses a Helium Neon laser as the source.

Results Section

The first part of the results section will be devoted to the construction and calibration of the visible light Mueller matrix polarimeter. The same design used to

construct the infrared polarimeter was implemented for the visible light polarimeter.

Figure 1 shows a block diagram of the polarimeter.

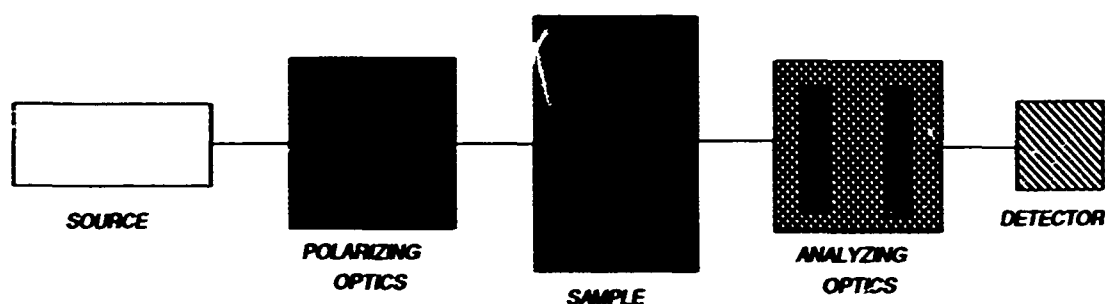


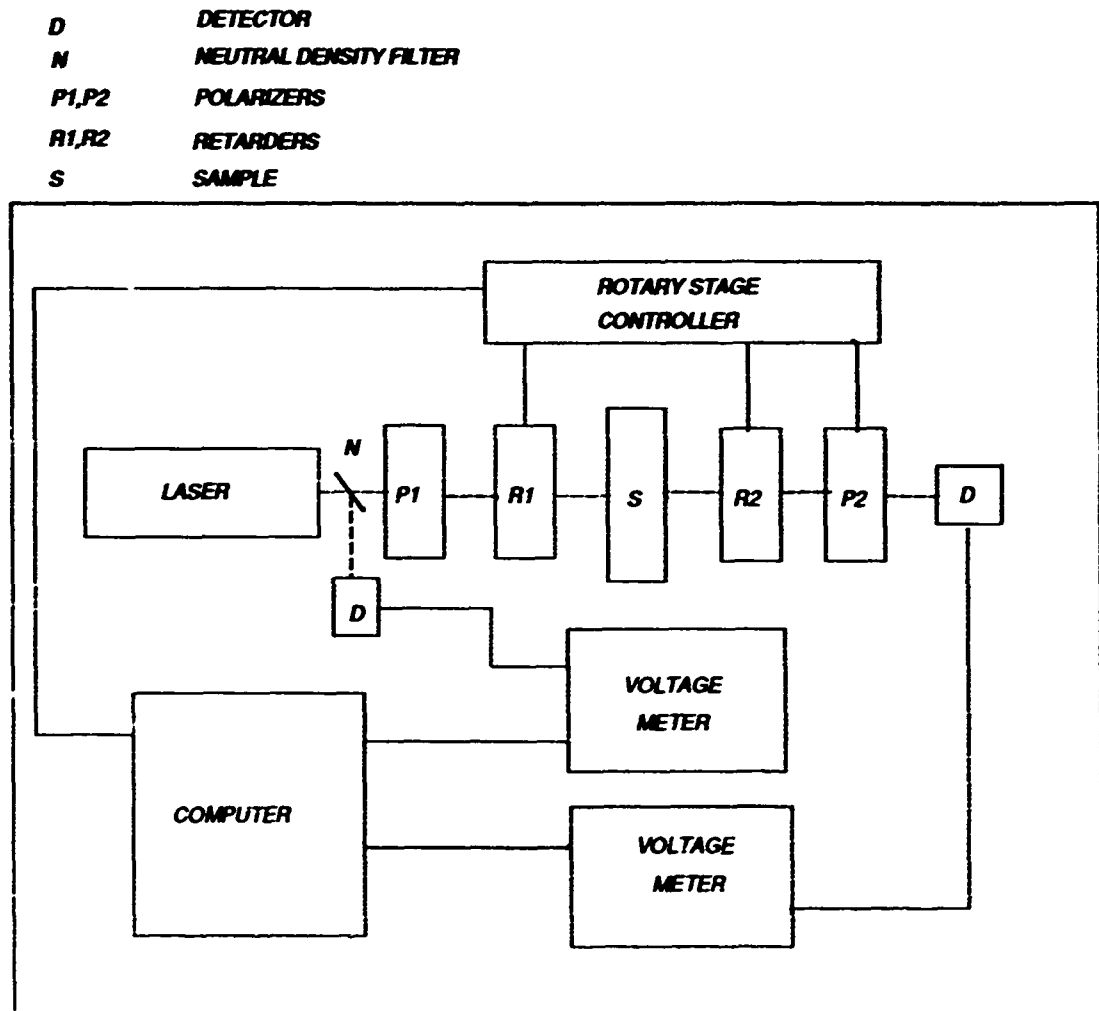
Figure 1: Block diagram for a Mueller matrix polarimeter

This design of polarimeter can be divided into five sections, the source, polarizing optics, sample, analyzing optics, and detector. The source can be polarized, partially polarized, or unpolarized. The polarizing optics consists of a fixed linear polarizer followed by a quarter wave retarder. The analyzing optics consists of a quarter wave retarder followed by a fixed linear polarizer. The detector can be polarization insensitive since the last element the light propagates through is the fixed linear polarizer. This design of polarimeter is typical of automatic rotating compensator Mueller matrix ellipsometers.

Figure 2 is a more detailed illustration of the polarimeter. The source is a Helium Neon laser operated at $633nm$. Light emitted from the laser first encounters a neutral density filter angled at 45 degrees. This filter is necessary; without it the detector following the analyzing optics would saturate. Also light reflected off the front surface of the filter is sent to a detector which monitors the lasers output before it is modulated. This second detector is used to determine the lasers stability while measurements are

being made. The neutral density filter chosen allows both detectors to operate in the linear portion of their response curve.

Figure 2: Helium Neon laser polarimetry set up.



The portion of the beam which is transmitted through the neutral density filter next encounters a Glan Thompson linear polarizer. The transmission axis of the polarizer is aligned along the direction of polarization of the beam emitted from the laser, thus making the beam very highly polarized. This polarizer would be unnecessary if the light emitted from the laser had a sufficiently high degree of polarization.

After the polarizer the beam is transmitted through a quarter wave plate, followed by the sample, followed by another quarter wave plate, followed by another Glan Thompson polarizer, then on to the detector. The two quarter wave plates and the final polarizer are all mounted in computer controlled rotary stages. The stages are connected to a programmable stage controller which is connected to the Hewlett Packard (HP) computer. The output of the detectors are sent to programmable multimeters which can be read by the HP computer.

The polarimeter is completely computer controlled with all data acquisition and data reduction being programmed in Basic. The computer is programmed to rotate the quarter wave plates and query the multimeters. The first wave plate is rotated through five degrees and the second through twenty five degrees. The stages are stopped after each incremental rotation and the data is collected from the multimeters. The first wave plate is rotated through a total of 360 degrees. The resulting data is a modulated intensity waveform which is processed according to an algorithm developed by Azzam.[5] The intensity waveform can be expressed as a Fourier series where the Fourier coefficients are functions of the sample Mueller matrix. These expressions can be inverted to give the Mueller matrix elements as functions of the Fourier coefficients. The Mueller matrix contains the polarization properties of the sample.

Several issues must be addressed in aligning and calibrating the polarimeter. It is important the fast axes of the retarders and the transmission axes of the second polarizer are aligned with the transmission axis of the first polarizer. Therefore, we define error terms for each of these orientations. Two other error terms are defined for the retarders being nonideal quarter wave plates. Figure 3 illustrates these error terms.

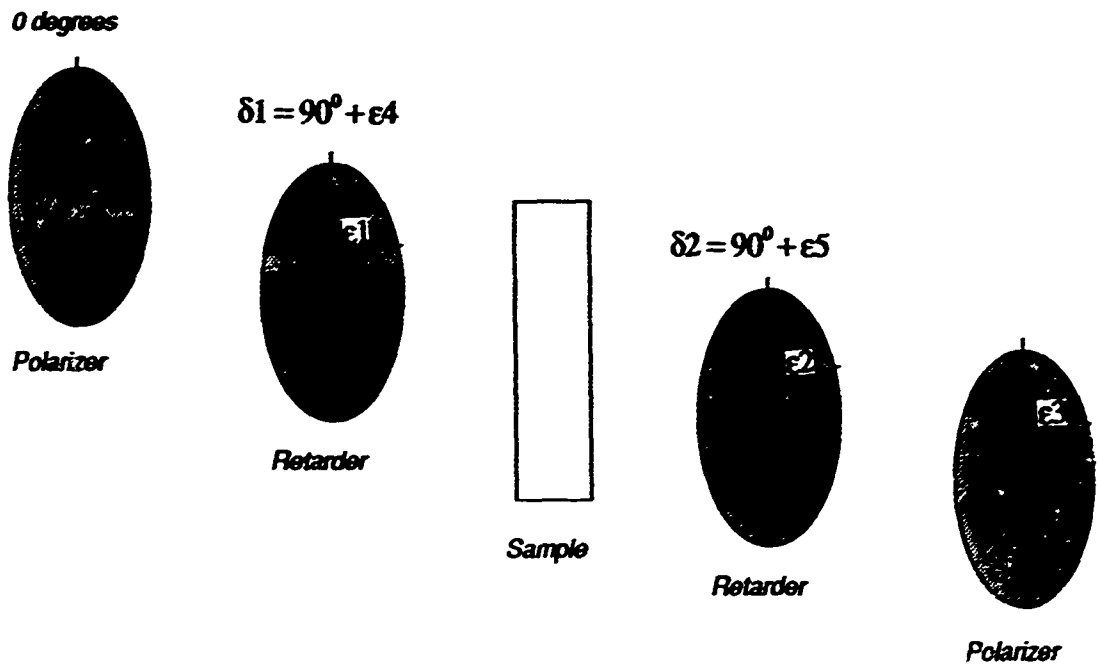


Figure 3: Error terms used for calibration and compensation.

The transmission axis of the first polarizer is defined as zero degrees of rotation. It is necessary to align the fast axes of the two retarders and the transmission axis of the final polarizer at zero degrees. A rough estimate of the orientation can be made from markings on the elements by the manufacturer, however, computer simulations have shown polarimeter measurements to be sensitive to these alignments, therefore a more accurate method of alignment is required.

By including the error terms $\epsilon 1, \epsilon 2, \epsilon 3, \delta 1$, and $\delta 2$ in the data reduction, the modulated intensity seen at the detector becomes a function of the Fourier coefficients, the sample Mueller matrix elements, and the error terms. This expression can be inverted to give the error terms as a function of the Fourier coefficients and the sample Mueller matrix elements.

If measurements are made with no sample (light propagating through air) the

optical properties of the beam will not be altered as it passes through the sample compartment. Therefore, the Mueller matrix for no sample is ideally the identity matrix. By substituting the identity matrix for the sample Mueller matrix the error term equations $\epsilon_1, \epsilon_2, \epsilon_3, \delta_1$ and δ_2 can be written solely in terms of the Fourier coefficients.

By running the measurement program with no sample in the sample compartment the Fourier coefficients can be generated from the modulated intensity waveform; from these the error terms can be calculated. The rotation stages, which are accurate to .001 degrees, can then be rotated the desired amount to correct the orientation errors. This process is repeated until the orientational error terms are driven as close to zero as possible. Generally, ϵ_1, ϵ_2 and ϵ_3 can be made smaller than .030 degrees after several runs with a stable laser.

Although the retardances of the quarter wave plates δ_1 and δ_2 can be determined in the calibration run, their retardances can not be changed since this is an inherent property of the quarter wave plates. However, the retardances along with the orientational error terms can be used to error compensate the sample Mueller matrix. This is done by solving the modulated intensity waveform for the Mueller matrix elements in terms of the Fourier coefficients and the error terms. A calibration run is made (one with no sample) before measuring a sample and the five error terms are calculated and recorded. Intensity measurements are then made with the sample in the sample compartment and the Fourier coefficients are determined from these. The error terms, generated from the calibration run, are entered into the program from the keyboard. The error compensated Mueller matrix is then calculated and displayed.

The following are Mueller matrices calculated with and without error compensation. Ideally these matrices would be the identity matrix. A comparison of the compensated and uncompensated matrices with the identity matrix shows the

considerable accuracy gained by using the error compensated equations. Although the compensated matrix is a vast improvement, there are still errors present. These errors may be caused by laser instabilities, detector fluctuations, and other error sources.

$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} .891 & -.042 & -.011 & 0.000 \\ -.058 & .999 & -.132 & -.001 \\ -.061 & .132 & 1.000 & .007 \\ 0.000 & .005 & -.012 & .891 \end{pmatrix}$	$\begin{pmatrix} .974 & -.004 & .004 & -.001 \\ -.004 & .999 & -.008 & -.001 \\ .000 & .008 & 1.000 & .007 \\ -.001 & .006 & -.011 & .973 \end{pmatrix}$
Identity matrix	Uncompensated matrix	Compensated matrix

The following results are for work on the infrared polarimeter. A new Edinburgh infrared laser was purchased during the past year. This laser can be tuned from $9\mu m$ to $11\mu m$. The power output of the laser depends on the wavelength to which the cavity is tuned. Therefore, it is desirable to know the power output for each of the lasers tunable wavelengths. A spectrum analyzer and an infrared detector were used to determine the power output for each of the discrete wavelengths from $9\mu m$ to $11\mu m$. Table 1 summarizes these measurements.

The first column is the micrometer drive setting on the laser, the second column is the wavelength, and the third column is the power output of the laser. Measurements could be made at nearly all of these wavelengths since the problem has generally been too much power on the detectors, not too little. Currently a beam splitter is used to dump 95% of the lasers output power to a graphite plate, this keeps the detectors from saturating. Polarimeter measurements could be made at wavelengths with lower power output by changing to a beamsplitter with a lower reflectivity. Currently measurements are made at $10.586\mu m$ since the retarders are designed to be quarter wave at $10.6\mu m$.

However, measurements made at a different wavelength would not require new retarders since the actual retardance is determined during calibration and the sample Mueller matrix is compensated with these values.

It was desirable to measure the polarization properties of a sample of CdTe as a function of electric field and determine its electrooptic coefficient. The application of an electric field can produce an alteration in a crystal index ellipsoid. This alteration can make the crystal anisotropic since its index of refraction may depend on direction. The phase retardation accumulated by polarized light in transversing a medium with anisotropic properties is given by

$$\delta = 2\pi(n_a - n_b)L/\lambda$$

where L is the medium thickness in the direction of propagation, λ is the wavelength of light, and n_a, n_b are the indices experienced in the two orthogonal directions perpendicular to the direction of propagation. If the light polarization and crystal are aligned so that the polarization is 45 degrees from either principal axis, the phase retardation will be

$$\delta = 2\pi(n'_y - n'_x)L/\lambda$$

where n'_y, n'_x are the new principal indices with the applied field. For a cubic crystal with $\bar{4}3m$ symmetry (such as CdTe) the principal indices are given by

$$n'_x = n_0 + \frac{1}{2}n_0^3 r_{41} E$$

$$n'_y = n_0 - \frac{1}{2}n_0^3 r_{41} E$$

where n_0 is the ordinary index of refraction, r_{41} is the electrooptic coefficient, and E is the electric field. If the electric field is expressed in terms of the electric potential and charge separation, then the phase retardation is

$$\delta = 2\pi n_0^3 r_{41} L / \lambda d$$

where d is the charge separation.

The Mueller matrix representation of a retarder with fast axes at arbitrary orientation angle θ is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\theta + \sin^2 2\theta \cos \delta & (1 - \cos \delta) \sin 2\theta \cos 2\theta & -\sin 2\theta \sin \delta \\ 0 & (1 - \cos \delta) \sin 2\theta \cos 2\theta & \sin^2 2\theta + \cos^2 2\theta \cos \delta & \cos 2\theta \sin \delta \\ 0 & \sin 2\theta \sin \delta & -\cos 2\theta \sin \delta & \cos \delta \end{pmatrix}$$

where δ is the retardance. If we assume the CdTe sample has negligible diattenuation and negligible circular retardance (an assumption that later will be shown to be valid for CdTe) then the Mueller matrices determined experimentally can also be represented by the linear retarder matrix. The linear retardance δ can be determined from several elements of the experimental Mueller matrix even if the fast axes orientation is unknown. The (4,4) element is independent of fast axes orientation, and the fast axes orientation can be eliminated elsewhere by adding the (2,2) and (3,3) matrix elements or squaring and adding elements in the fourth row and column. Therefore, given a measured Mueller matrix, a known applied voltage, and a known refractive index the, electrooptic coefficient r_{41} can be determined.[6]

Infrared Mueller matrix polarimeter measurements were made on a sample of CdTe. Mueller matrices for CdTe were determined as a function of an applied voltage across the crystal, from 0 to 3000 volts. The sample is a cubic crystal with dimensions of

49.6mm by 4.96mm by 4.96mm. The laser beam propagated through the 49.6mm length of the crystal which was oriented in the $\langle 110 \rangle$ plane. The sample was placed between two strips of indium which were used as conducting surfaces, this arrangement is situated between two metal plates. The metal plates were connected to a high voltage power supply and a 1000 to 1 attenuation probe was used to read the voltage from a digital multimeter.

A linear regression was done for six data sets consisting of inverse trigometric functions of experimental Mueller matrix elements versus expressions for the retardance as a function of voltage. The resultant slopes are then the values of $n_0^3 r_{41}$. The data sets consist of the inverse cosines of element (4,4) and the sum of element (2,2) and (3,3), and the inverse sines of the element combinations $\sqrt{(4,2)^2 + (4,3)^2}$, $\sqrt{(4,2)^2 + (3,4)^2}$, $\sqrt{(2,4)^2 + (4,3)^2}$, and $\sqrt{(2,4)^2 + (3,4)^2}$. The results of the regression analysis are given in Table 2.

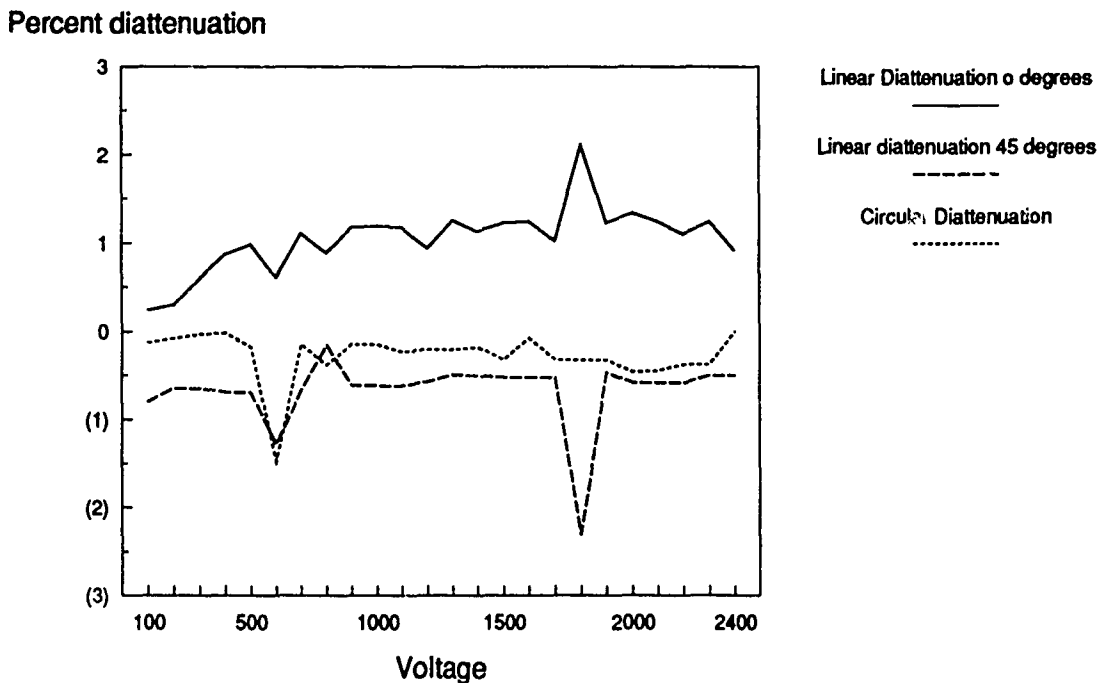
The average value of $n_0^3 r_{41}$ is 1.072×10^{-10} which corresponds to a value of $r_{41} = 6.10 \times 10^{-12}$ for the electrooptic coefficient, using an index of refraction of 2.6 for CdTe at $10.6 \mu m$.

A method for obtaining the electrooptic coefficient from the measured Mueller matrices has been shown. Although this is an important result for the study of electrooptic modulator materials, there may be other properties of the material which can be extracted from the Mueller matrix. A method has been established by Cloude which determines linear retardance and linear diattenuation along the x and y coordinate axes, linear retardance and linear diattenuation along plus and minus 45 degrees, circular retardance and circular diattenuation.

Cloude's method originates from group theory, where the Mueller matrix is written

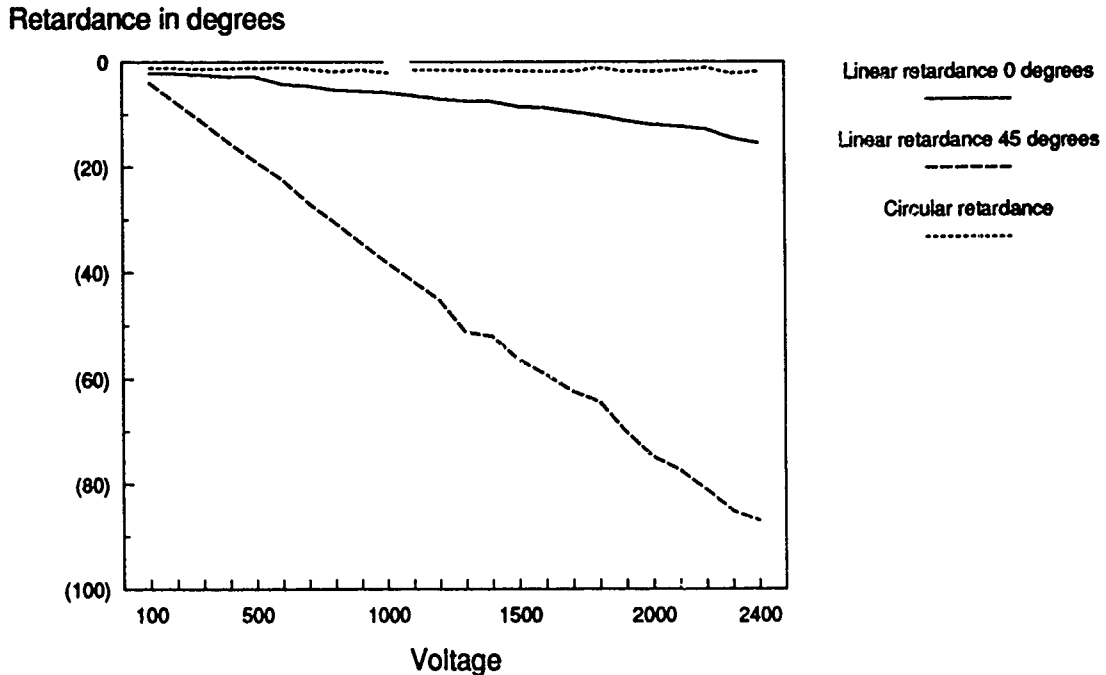
in terms of a 4 X 4 coherency matrix. The eigenvalues and eigenvectors are determined for this matrix, the dominant eigenvalue and eigenvector describes the dominant polarizing effect the material exhibits. Cloude's method was applied to the measured CdTe Mueller matrices. Plots of the diattenuation and retardance appear in Figures 4 and Figure 5.

Figure 4 Plot of diattenuation vs. Voltage



All three diattenuation terms were determined to be less than 1.5% and the circular retardance term was determined to be less than 2 degrees, for the voltage range of 100 to 2400 volts. Neither the diattenuation or the circular retardance was found to depend on voltage. The linear retardance terms were found to increase linearly with voltage from 100 to 2400 volts. These results show CdTe behaves as a linear retarder, with its

Figure 5 Retardance vs. Voltage



retardance being determined by the voltage applied across the crystal. Neglecting diattenuation and circular retardance terms (as was done earlier) will not effect calculation of the electrooptic coefficient significantly.

Conclusions and Recommendations

The method established for compensating the Mueller matrices for element orientation errors and nonideal retardance errors has been shown to improve accuracy in measurements made on both the infrared and visible light polarimeters. This method and the addition of the new Edinburgh infrared laser has allowed for accurate measurements to be made on a sample of CdTe. These results will be submitted to Applied Optics. Construction of the Helium Neon polarimeter will allow for Mueller matrix measurements to be made at visible wavelengths. The Helium Neon polarimeter is

currently accurate to about 3%; with further alignment and calibration work the polarimeter could be made accurate to 1%.

I would like to offer some recommendations for improving both the visible and infrared polarimeters. The visible polarimeter is now operational; however, more calibration work is needed to improve its accuracy to a satisfactory level. Currently the two quarter wave retarders are mounted in holders which allows the retarders to be rotated so the face of the retarders are perpendicular to the incident beam. These holders are not as stable as desirable, making them subject to vibrations. Without these holders the incident beam would strike the retarders at a small angle from normal and beam wander may result as the retarders are rotated during the measurement process. This would be a problem if the beam wandered off the face of the detector, however this problem could be solved with the use of an integrating sphere in front of the detector. Other alignment work may be necessary to improve the visible polarimeter performance.

An idea that may improve the results of both the visible and infrared polarimeters, is to compensate the data for power fluctuations from the laser. Both polarimeters use a detector for taking data and one for monitoring the lasers power output. Therefore, a laser power output measurement is made at the same time each of the discrete elements of data are taken. This information should allow the sample data to be compensated for power fluctuations from the laser.

Whether the previously mentioned plan is implemented or not, it would be a good idea to install an integrating sphere in front of the infrared laser power monitoring detector. The integrating sphere will attenuate the beam allowing the detector to operate without being saturated. Currently several beam splitters are used to divert portions of the beam, however, the detector is still nearly saturated. The integrating sphere will also aid in aligning the beam onto the detecting surface since the port on the integrating

sphere is much larger than the surface of the detector.

The construction of a solenoid, for placing samples in a uniform magnetic field, will allow for Faraday rotation measurements to be made as a function of magnetic field. This would mainly be used in the infrared polarimeter for characterizing potential magneto optic modulator materials.

It may be of interest in the future to measure the polarization characteristics of optical waveguides and fiber optic cables. This would require additional elements in the polarimeter for focusing the beam into the waveguides and cables. The polarimeter would have to be recalibrated with these elements in place. Whether it would be difficult to calibrate the polarimeter with these elements or whether it is possible to keep the beam focused on the waveguide or cable while the retarders are rotating are areas for study. Further work in this area may point to a new class of samples which could be measured.

Acknowledgements

First of all I would like to thank AFOSR for sponsoring the research program and RDL for overseeing and simplifying the paper work. I have enjoyed working for my focal point Dr. Dennis Goldstein and would like to thank him for allowing me to work on such interesting projects and offering sound advice. Thanks is also extended to members of the special projects lab Howard McCormick, Mike VanTassel, Voncile Ashley, Linda Lau, and David Onuffer for their quick and practical solutions to every problem I encountered. A special thanks goes to Annette Marsh and Chad Houghton for their contributions to this summers experiences.

Table 1

Drive Reading	Wavelength Microns	Output Power Watts	Drive Reading	Wavelength	Output Power Watts
235.8	9.190	3.5	456.7	10.153	1.3
238.5	9.210	4.5	458.2	10.163	2.1
240.2	9.218	4.9	463.2	10.178	3.4
241.1	9.220	5.7	465.8	10.190	4.4
243.2	9.230	6.5	469.2	10.202	4.5
244.8	9.237	7.0	471.5	10.219	4.7
247.3	9.245	7.2	472.0	10.230	4.8
250.6	9.255	6.8	478.4	10.242	4.8
254.8	9.278	6.8	479.7	10.256	5.5
258.0	9.290	6.7	482.8	10.268	5.6
261.4	9.304	6.0	486.5	10.283	5.4
263.6	9.313	5.9	488.6	10.300	5.4
266.5	9.323	4.7	490.9	10.308	5.3
268.5	9.340	3.3	495.1	10.312	4.5
296.5	9.458	2.5	529.2	10.490	2.7
300.8	9.470	4.6	533.0	10.510	4.3
303.4	9.490	5.3	538.3	10.523	4.8
308.5	9.503	5.4	543.4	10.543	4.9
312.3	9.520	6.0	547.5	10.565	5.4
317.0	9.537	6.0	551.0	10.582	5.7
321.8	9.550	6.5	555.9	10.586	5.7
325.2	9.568	6.7	560.4	10.602	5.6
328.2	9.584	6.5	566.2	10.623	5.3
333.0	9.602	6.4	571.0	10.643	4.8
338.3	9.620	5.6	576.0	10.690	4.0
340.1	9.639	4.7	578.4	10.712	3.6
345.1	9.658	4.6	585.0	10.738	3.0
348.0	9.678	4.1	590.3	10.760	2.4
353.3	9.693	3.8	594.5	10.783	1.5
356.2	9.715	3.0			
362.6	9.735	2.0			

Table 2

Mueller matrix elements	Slope = $n_0^3 r_{41}$	Standard deviation
(4,4)	1.095×10^{-10}	.0084
(2,2),(3,3)	1.077×10^{-10}	.0113
(4,2),(4,3)	1.056×10^{-10}	.0078
(2,4),(3,4)	1.060×10^{-10}	.0162
(4,2),(3,4)	1.096×10^{-10}	.0105
(2,4),(4,3)	1.065×10^{-10}	.0072
Average	1.072×10^{-10}	2.2×10^{-12}

Endnotes

1. Chipman, Russell A., "Polarized light and polarimetry", class notes, volume 1, pg. 209.
2. Chenault, David B., "Mueller Matrix Infrared Polarimetry", Final Report USAF-UES Graduate student research program, August 1988.
3. Hodgson, Randall R., "Laser Polarimetry Development", Final Report USAF-UES Graduate student research program, August 1989.
4. Gove Randy L., "Infrared Laser Polarimetry", Final Report USAF-UES Graduate student research program, August 1990.
5. Azzam, R.M.A., "Photopolarimetric measurement of the Mueller matrix by Fourier analysis of a single detected signal," Opt. Lett., 2, 148-150, 1978
6. Goldstein, Dennis H., "Polarization Modulation in Infrared Electrooptic Modulator Materials", Dissertation, University of Alabama in Huntsville, 1990

MATERIAL CHARACTERIZATION AND EVALUATION FOR PENETRATOR APPLICATIONS

Graduate Student Steve Hatfield

As a participant in the Air Force Office of Scientific Research (AFOSR) Summer Research Program for graduate students, I learned the importance of material selection for dynamic penetration mechanics. The selection process to determine the optimum material involved the operation and interpretation of several precise tests. These tests consist of the Hydraulic Tension Test, the Hopkinson Bar Test and the Rod Impact Test, most commonly known as the Classic Taylor Test. In addition to operating these test the lab was also evaluating material properties for other clientele. I was also involved in upgrading the computer program that the Hydraulic Tension Test used to achieve useable data and produce stress-strain plots. I am hoping to continue the research this summer at the University of Kentucky.

The main focus of the summer research was the material characterization and evaluation for penetrator applications. The question to be answered was if the material properties used in the WL/MNV lab experimentation were equivalent to the material properties used by Los Alamos National Laboratory, (LANL), in their MTS computer codes. The material tested was oxygen free electron copper (OFE Cu). The lab was checking through experimentation the results of the MTS model compared to actual test results. The MTS codes are used to more effectively determine how a material will react upon impact. These codes are a valuable tool in dynamic penetration mechanics. The codes can save hundreds of research hours involved with the fabrication of warheads.

The approach taken to solve this problem was an evaluation of the material characteristics involving comparison, testing and results. This was followed by controlled testing which paralleled the MTS codes. And finally post impact analysis which gave us the finally results. The post impact analysis was produced by Jessica Mayes, a fellow graduate student.

The initial material characteristics were controlled through the series of tests discussed earlier.

The comparison of LANL and model shop OFE Cu was needed because the MTS model used OFE Cu annealed at 600 C for 1 hour. This comparison of material properties was established by the engineering tension test. The comparison consisted of LANL sending some material samples that were annealed in the model shop furnace. This material was tested, then the model shop took in house material which was the same and annealed it under the same conditions. The data was then collected and evaluated. Before the samples were annealed they was machined into ASTM standard tensile test specimen (see figure 1).

The material properties that we were most interested in were the yield strength of the material, the ultimate tensile strength of the material and the ductility or the total elongation of the material. These properties of all the materials were checked against each other to look for similarity. A table of the material tested is given in table 1.

The properties were evaluated by examining the model shop OFE copper as received and annealed for 1 hour at 600 C, then examining the LANL OFE copper as received, 600 F and 600 C. The results were then compared against one another to determine if the two coppers

exhibit the same history. The model shop OFE copper's history changed very little with annealing. The yield strength of the as received was 12.0 ksi while the yield strength of the annealed was 4.2 ksi. This difference is attributed to stress relaxation taking place due to the annealing at 600 C for 1 hour. Looking at the ultimate tensile strengths they were about the same 30.0 ksi for the as received and 31.0 ksi for the annealed. The total elongation was 82.3 % for the as received and 69.3 % for the annealed. All data is documented in table 2 and table 3. The stress-strain plot of the model shop OFE copper, as received, is given in figure 2 while the stress-strain plot for the annealed model shop OFE copper is figure 3.

The same procedure was done for the LANL OFE copper except the model shop made an error and annealed the first sample at 600 F instead of 600 C. This error was revealed through the evaluation of the collected data. As can be seen looking at figures 4,5 and 6 the stress-strain curves are significantly different at each stage. The differences in the as received material is obvious. Comparing figure 2 and 4 we see a drastic change from one material to the other. This change is because the LANL OFE copper, as received, is

half hard. This means that the LANL copper has been cold worked by extrusion into rod form. This cold working caused the change in material properties of the two as received coppers. The cold working is what also causes the change in material properties through stress relaxation in the LANL as received, 600 F and 600 C. As can be seen in figures 4,5 and 6 the higher the temperature of annealing, or increase in time, the more effect that stress relaxation has on a material. A comparison of the yield strengths, ultimate tensile strengths and the total elongations of LANL copper can be made by comparing table 2 and 3.

The reason for all the testing was to determine if the material used in the MTS computer codes and the actual material used for penetration mechanics exhibited the same mechanical properties. The testing was done to determine if the model shop could attain the same mechanical properties of annealing that were attained by LANL. The testing showed that the material properties in the LANL MTS model matched the properties of the material that the model shop had in stock and that penetrators were being fabricated from. The test also showed that the model shop could achieve the same properties through annealing that were attained by LANL. Although this report is short

in length, it does not reflect the numerous weeks spent learning the operation of the testing equipment that made this report possible.

Materials Evaluated:

OFE Cu LANL (As Received)
OFE Cu Model Shop (As Received)

OFE Cu LANL (600F, 1hr.)
OFE Cu LANL (600C, 1hr.)
OFE Cu Model Shop (600C, 1hr.)

NOTE: LANL Represents Los Alamos National Laboratory

Table 1

Comparison of Stress-Strain Data Between LANL and Model Shop OFE Cu

Sample	Yield Strength	Ultimate Tensile Str	Total Elongation
● LANL As Received	10.2 ksi	50.0 ksi	15.9 %
● LANL As Received	14.4 ksi	49.8 ksi	14.4 %
● Model S (Axial) As Received	12.0 ksi	31.0 ksi	69.3 %
● Model Shop (Transverse) As Received	9.5 ksi	29.5 ksi	78.2 %

NOTE: Engineering Strengths Listed

Table 2

Comparison of Stress-Strain Data Between LANL and Model Shop OFE Cu

Sample	Yield Strength	Ultimate Tensile Str	Total Elongation
● LANL (600F, 1 hr.)	14.0 ksi	43.5 ksi	25.7 %
● LANL (600C, 1 hr.)	3.6 ksi	31.8 ksi	81.8 %
● Model Shop (Axial) (600C, 1 hr.)	4.2 ksi	30.0 ksi	82.3 %
● Model Shop (Transverse) (600C, 1hr.)	4.2 ksi	29.4 ksi	79.5 %

NOTE: Engineering Strengths Listed

Table 3

ASTM Standard Test Specimen For Tensile Test

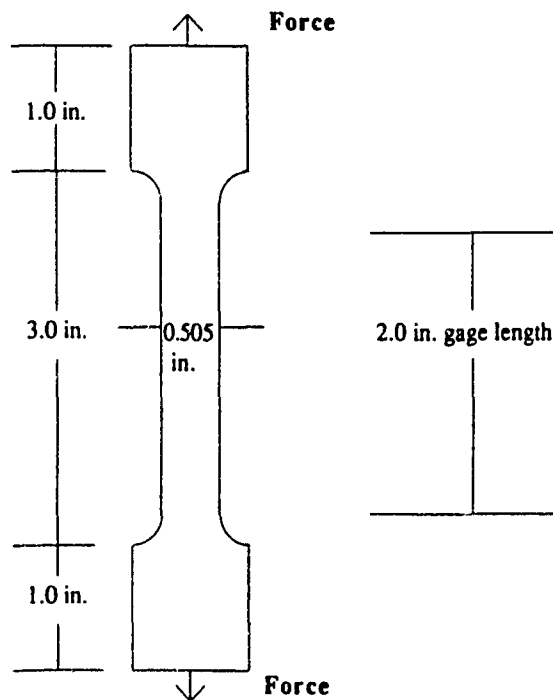


Figure 1: ASTM Tensile specimen

ENGINEERING STRESS-STRAIN CURVE
 TEST NUMBER: INSTRON #299
 DATE: 18 JUNE 91
 OFE CU RCVD L

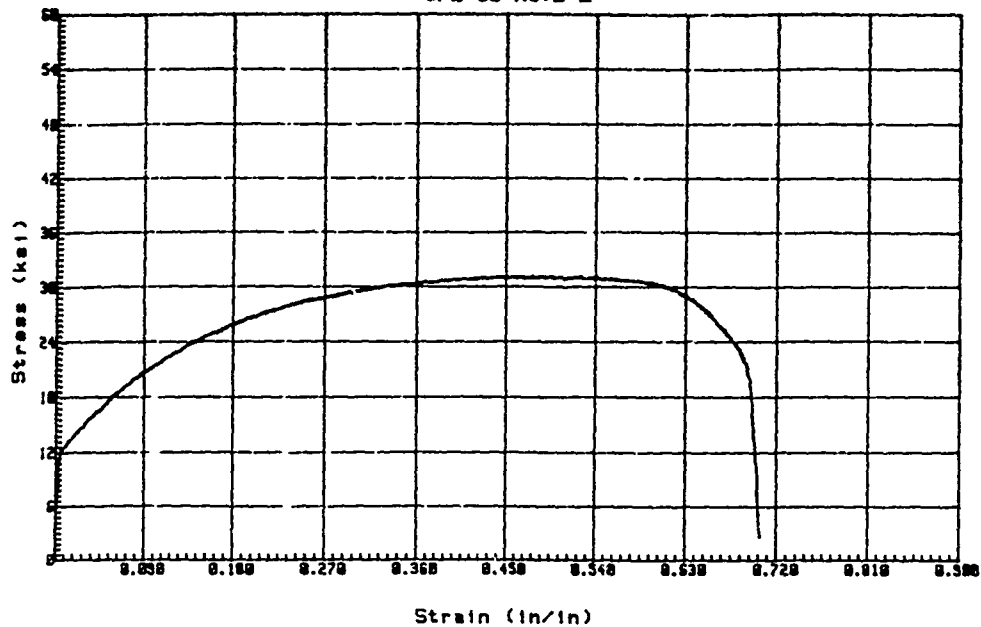


Figure 2: Model Shop OFE copper (As Received)

ENGINEERING STRESS-STRAIN CURVE
 TEST NUMBER: INSTRON #297
 DATE: 17 JUNE 91
 ANNEALED OFE CU RX

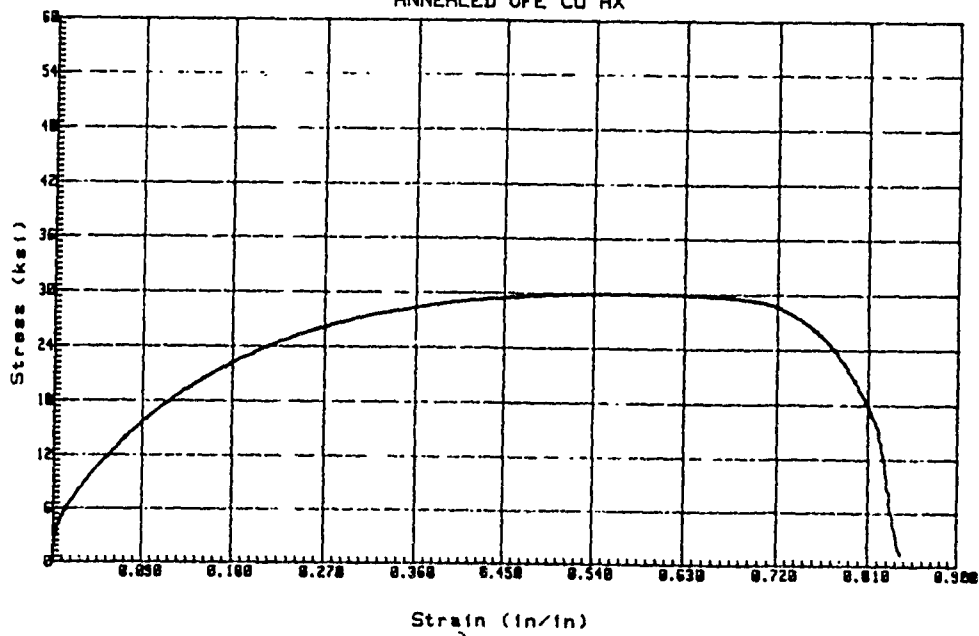


Figure 3: Model Shop OFE copper (annealed 600 C, 1 hour)

ENGINEERING STRESS-STRAIN CURVE
 TEST NUMBER: INSTRON 0296
 DATE: 13 JUNE 91
 OFE CU LANL

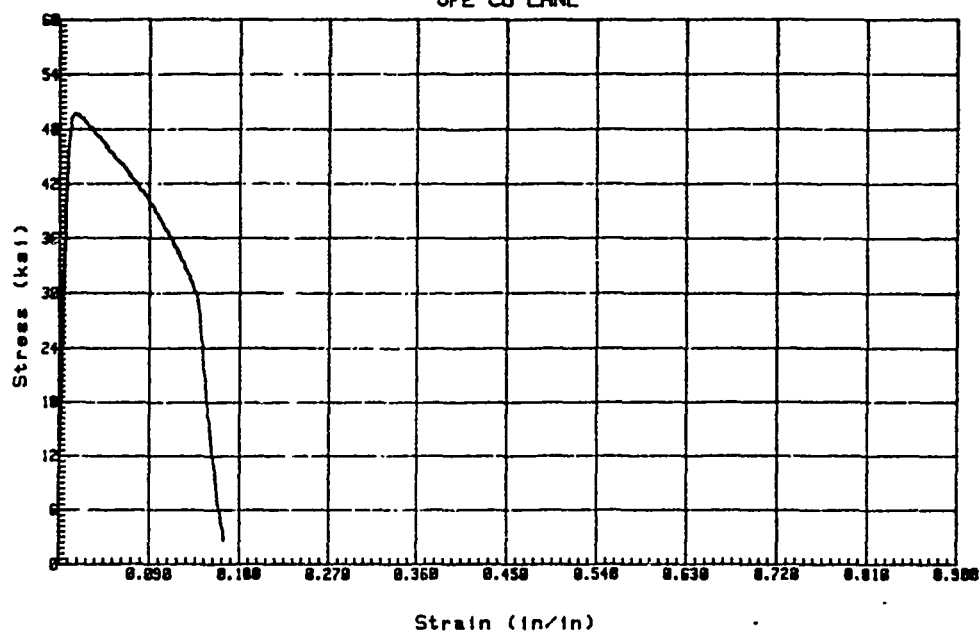


Figure 4: LANL OFE copper (As Received)

ENGINEERING STRESS-STRAIN CURVE
 TEST NUMBER: 301
 DATE: 1 JULY 91
 OFE CU LANL 600F

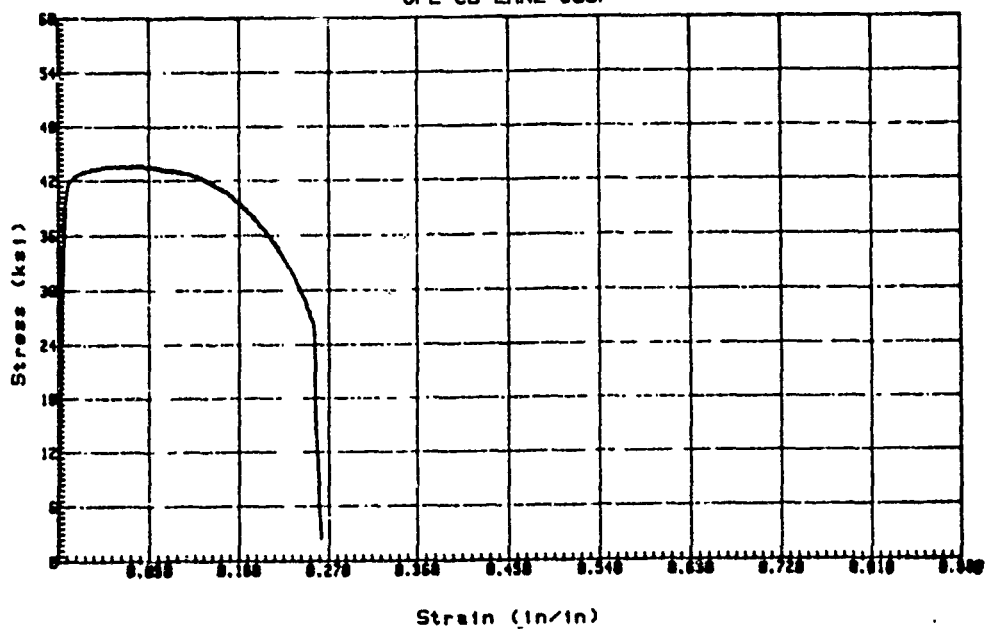


Figure 5: LANL OFE copper (annealed 600 F, 1 hour)

ENGINEERING STRESS-STRAIN CURVE
TEST NUMBER: INSTRON #303
DATE: 15 JULY 91
LANL CU 600C 1HR

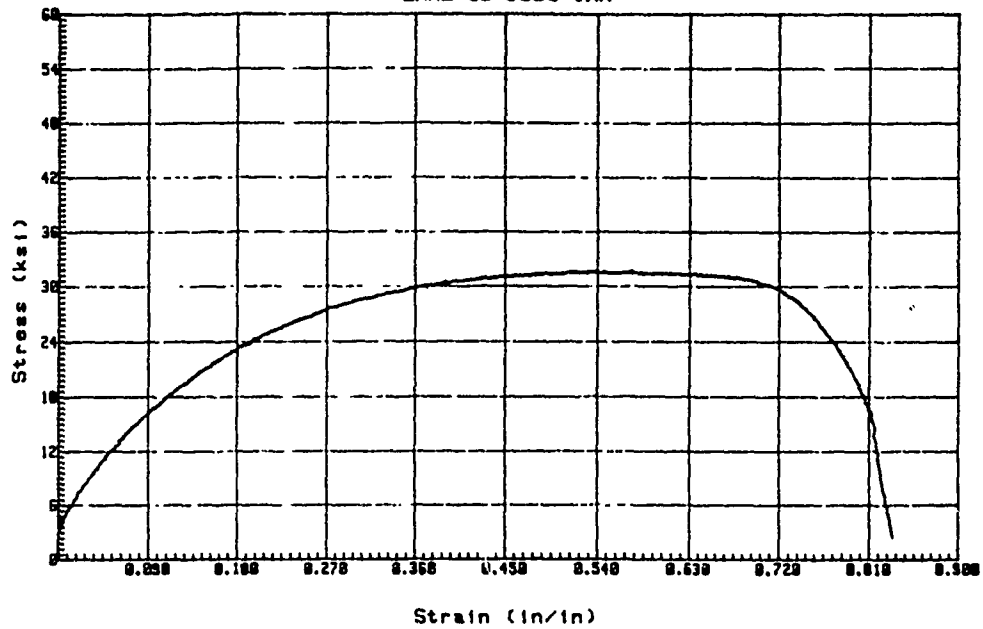


Figure 6: LANL OFE copper (annealed 600 C, 1 hour)

TWO-DIMENSIONAL WAVEGUIDE SIMULATION

Christopher P. Hussell

ABSTRACT

A program was developed to implement the spectral index technique for calculating the mode indices and field profiles of rib waveguide structures and the results were compared to those previously published. The method was extended for the case of dual rib waveguiding structures and a new program written. Since accurate data on structures of this type are scarce, an attempt was made to develop a finite element solution to the waveguide structure using a full vector field. This would be used to verify the accuracy of the extensions made to the spectral index technique. This remains to be completed.

ACKNOWLEDGMENTS

I would like to thank the Air Force Office of Scientific Research for sponsoring this important program and Research and Development Laboratories for their mediation between the Air Force and the research associates. I found this system to be efficient and the funds were dispersed in a timely manner.

My greatest appreciation goes to my focal point, Dr. Dennis Goldstein. He was very supportive and provided a conducive environment for research. I appreciate his efforts in recruiting applicants, providing information on housing, and warmly introducing me to the base. He made sure that I had everything I needed and provided many helpful discussions pertinent to my research.

Special appreciation is extended to Dr. Tom Davis for helping me to familiarize myself with the base computers and suggestions regarding signal processing, and Dr. Charles Martin who gave me suggestions regarding finite elements. Mac McCormick, David Onuffer, and Linda Lau should be commended for their technical support and providing me with a well-equipped and well-organized optical laboratory. Others whom I would like to thank are Dr. Gene Chenette, Lynn Deibler, Randy Gove, Annette Marsh, Lisa Collins, Danielle Walker, Chad Houghton, and David Gray for helpful discussions and providing a friendly atmosphere.

I. INTRODUCTION

Rib waveguiding structures have become an important technology as a result of recent advances in the growth of III-V semiconductor epitaxial layers. These structures may be incorporated onto the same substrate as other optical and electrical components performing such functions as connecting waveguides, bends, tapers, power dividers, switches, modulators, filters, and multiplexers/demultiplexers. Such a broad spectrum of applications for the rib waveguide justifies the search for the accurate knowledge of their propagation constants and the field profiles which are essential for the design of new devices.

Several factors contribute to making this a difficult endeavor. First, exact analytical solutions to rectangular waveguides bounded by dielectric materials are not available. This stems from the fact that the solution to the scalar wave equation in a rectangular medium with the boundary conditions imposed by a dielectric—dielectric interface cannot be expressed in terms of known analytic functions. Second, numerical solutions of the scalar wave equation are insufficient across the boundaries of most waveguide structures in semiconductor materials because they are only accurate when the gradient of the refractive index profile is small. Furthermore, the true modes of rib waveguides have electric and magnetic field components in both the transverse and longitudinal directions and the modes may only be quasi-TE or quasi-TM. Therefore, the exact treatment requires the consideration of all field components of the vector-wave equation. Rahman and Davies [1,2,3] formulated such a vector solution for arbitrary refractive index profiles using a finite element technique. Their results have been used recently as a benchmark for accuracy [4]. This formulation, however, is very numerically intensive and may not be efficiently incorporated into a mode propagation program such as that of Burns and Milton [5,6] to analyze rib waveguide devices. Therefore it is often necessary to resort to approximate methods that yield accurate field profiles and propagation constants.

The spectral index method, developed by McIlroy et al. [4], was shown to give accurate propagation constants as well as field profiles for a single rib waveguide. This method is attractive since it requires only the solution of a simple dispersion relation for the propagation constant and a one-dimensional Fourier transform for the field profile. However, the method is applicable to only specific cases, namely a single rib waveguide structure with high index difference between the film and the cladding. Since this is a very important structure in semiconductor devices, its study is justified.

The first portion of this report presents an extension to the spectral index method to obtain field profiles and propagation constants of the normal modes of multiple waveguide structures. This information allows us to immediately calculate the characteristics of symmetric and asymmetric directional couplers [7] and, using a local normal mode propagation technique such as that used by Hussell et al. [6], many multiple waveguide devices may be accurately and efficiently simulated. Several characteristics of the spectral index method make it a favorable method for these purposes: 1) it yields accurate field profiles as well as propagation constants, 2) the propagation constants are found via a simple dispersion relation thus not requiring a solution of the eigenvalues of a large matrix, 3) the dispersion relation may be differentiated with respect to the propagation constant allowing the use of fast iteration techniques such as Newton's method, and 4) only a one-dimensional inverse Fourier transform is required to obtain the field profiles from their respective propagation constants.

Since there is little data available to compare the solution we obtain from the modified spectral index method, a program must be developed to solve the same waveguide structure by a proven method to verify its accuracy. The second part of this report briefly explains the finite element method which will be used to make such a comparison. The finite element method will serve very well as a benchmark for accuracy since only a few comparisons need to be made. However, since it requires the solution of the eigenvalues of a very large matrix, it would be impractical to incorporate it in a device simulation program.

II. SPECTRAL INDEX METHOD

Figure 1 shows the dual waveguide structure of interest. The ribs are separated by $2S$, rib widths of W_1 and W_2 , heights of H_1 and H_2 , and a guiding film thickness of D . The refractive index of each region is also shown in the figure. Region 1 is the cladding (usually air), region 2 is the guiding layer, region 3 is the substrate, and regions 4 and 5 are strips of different indices. The principle of the spectral-index technique is to express the solution in the regions below the rib in terms of a Fourier transform in the x direction and approximate the solution in the rib in terms of sines and cosines with a field of zero in the cladding. These two solutions are linked by requiring continuity and using Parseval's formula to obtain a

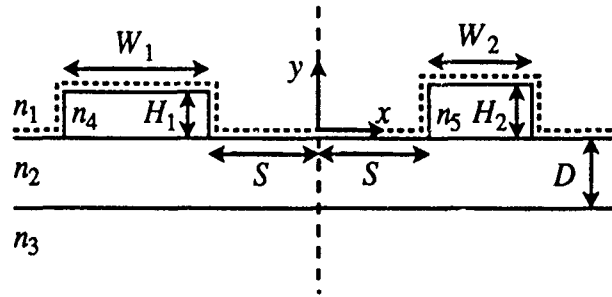


Figure 1. Dual rib waveguide structure.

simple dispersion relation. The method is outlined as follows.

The solution of the field in each region is a solution of the scalar wave equation,

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + (k^2 - \beta^2)E = 0,$$

with proper boundary conditions imposed. If the refractive index of the cladding is much smaller than the other regions (the case when air or an oxide is used), the field may be approximated to be zero on the interface if the boundary is moved using the Goos-Hänchen shift. These new boundaries are shown by the dotted line in Figure 1 and the new dimensions are given by

$$W'_{1,2} = W_{1,2} + \frac{2k_1^2}{p k_{4,5}^2 \sqrt{\beta^2 - k_1^2}},$$

$$D' = D + \frac{q}{\sqrt{\beta^2 - k_1^2}},$$

and

$$H'_{1,2} = H_{1,2} + \frac{p - q}{\sqrt{\beta^2 - k_1^2}}$$

where for TE modes $p = q = 1$ and for TM modes $p = k_1^2/k_{4,5}^2$ and $q = k_1^2/k_2^2$.

Using a Fourier transform with respect to x on the wave equation, we have

$$\frac{\partial^2 \mathcal{E}}{\partial y^2} + (k^2(y) - s^2 - \beta^2) \mathcal{E} = 0$$

where $\mathcal{E}(s, y)$ is the Fourier transform of $E(x, y)$. The solution is similar to that of a slab waveguide with a spectral index $n_s = \sqrt{n^2 - s^2/k_0^2}$.

The solution in the rib regions may be expressed as $E(x, y) = F(x)G(y)$ where

$$F(x) = \begin{cases} A_1 \cos(s_1(x+C_1)) & -C_1-W'_1/2 < x < -C_1+W'_1/2 \\ \pm A_2 \cos(s_2(x-C_2)) & C_2-W'_2/2 < x < C_2+W'_2/2 \\ 0 & \text{elsewhere} \end{cases}$$

$$G(y) = \begin{cases} \frac{\sin\{\gamma_1(-y+H'_1)\}}{\sin(\gamma_1 H'_1)} & -C_1-W'_1/2 < x < -C_1+W'_1/2 \\ \frac{\sin\{\gamma_2(-y+H'_2)\}}{\sin(\gamma_2 H'_2)} & C_2-W'_2/2 < x < C_2+W'_2/2 \\ 0 & \text{elsewhere} \end{cases}$$

$$C_1 = S + W_1/2, C_2 = S + W_2/2, s_1 = \pi/W_1', s_2 = \pi/W_2',$$

$$\gamma_1 = \sqrt{k_4^2 - s_1^2 - \beta^2}, \gamma_2 = \sqrt{k_5^2 - s_2^2 - \beta^2},$$

where k are the propagation constants of the respective regions, β is the propagation constant of the mode and the + represents the fundamental mode while the - represents the first higher-order mode.

Using Parseval's formula it may be shown that the spectral solution below the rib and the real-space solution within the rib may be linked by [4]

$$\int_{-\infty}^{\infty} E \frac{\partial E^*}{\partial y} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{E} \frac{\partial \mathcal{E}^*}{\partial y} ds$$

This represents a dispersion relation of which there are three unknowns, A_1 , A_2 , and β . One of the field amplitudes is arbitrary while the other may be used to determine the correct field profile. Since β depends on the field profile variationally, an extremum in β will be observed when the correct value for A_1/A_2 is used. Thus, the dispersion relation above must be used in conjunction with an optimization routine to solve multiple waveguide structures.

A computer program was developed to implement the procedure outlined above. The program was written in standard FORTRAN and run on a Macintosh SE/30 computer. The typical run time required to determine the propagation constants of a dual waveguide structure was less than four minutes. The program was tested using typical values for a rib waveguide structure in a GaAs/AlGaAs system. For the results given in this report, the following parameters were used for the TE mode: $W_1 = 3\mu\text{m}$, $W_2 = 3\mu\text{m}$, $D = 0.5\mu\text{m}$, $H_1 = 0.5\mu\text{m}$, $H_2 = 0.55\mu\text{m}$, $S = 2.0\mu\text{m}$, $n_1 = 1$, $n_2 = 3.44$, $n_3 = 3.4$, $n_4 = 3.445$, $n_5 = n_2$, and the

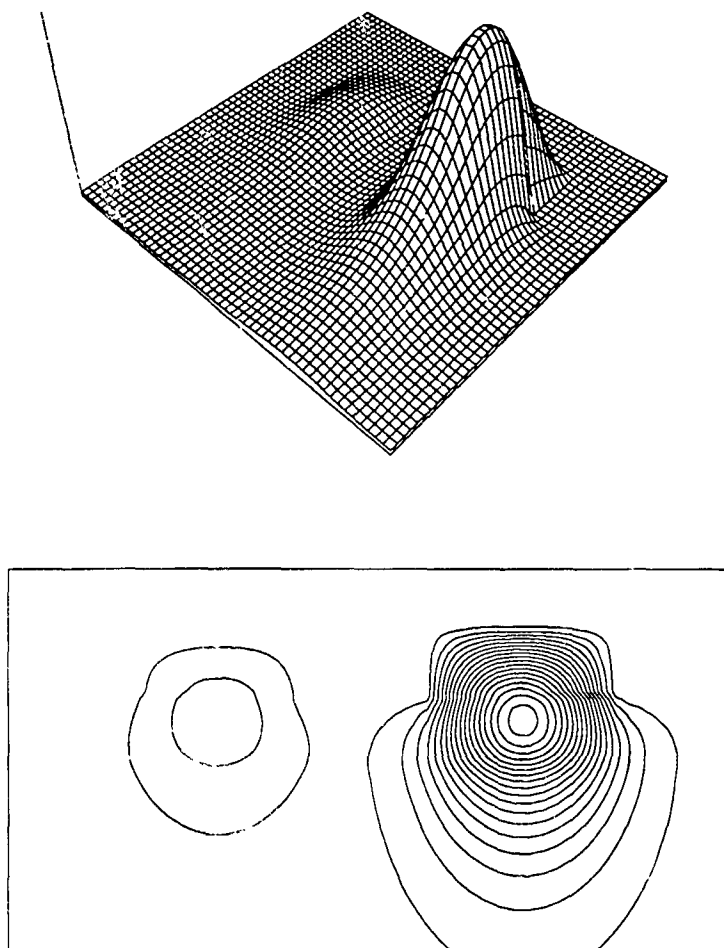


Figure 2. Surface and contour plots of the fundamental field profile of the dual rib waveguide structure. The parameters are given in the text.

wavelength $\lambda = 1.15\mu\text{m}$. The surface and contour plots of the fundamental and first order modes are shown in Figures 2 and 3 respectively. Their corresponding normalized propagation constants were found to be 0.35889 and 0.35044 respectively. By isolating each waveguide and calculating their propagation constants using the method of McIlroy et al. [4], the normalized propagation constants of their fundamental modes were found to be 0.35056

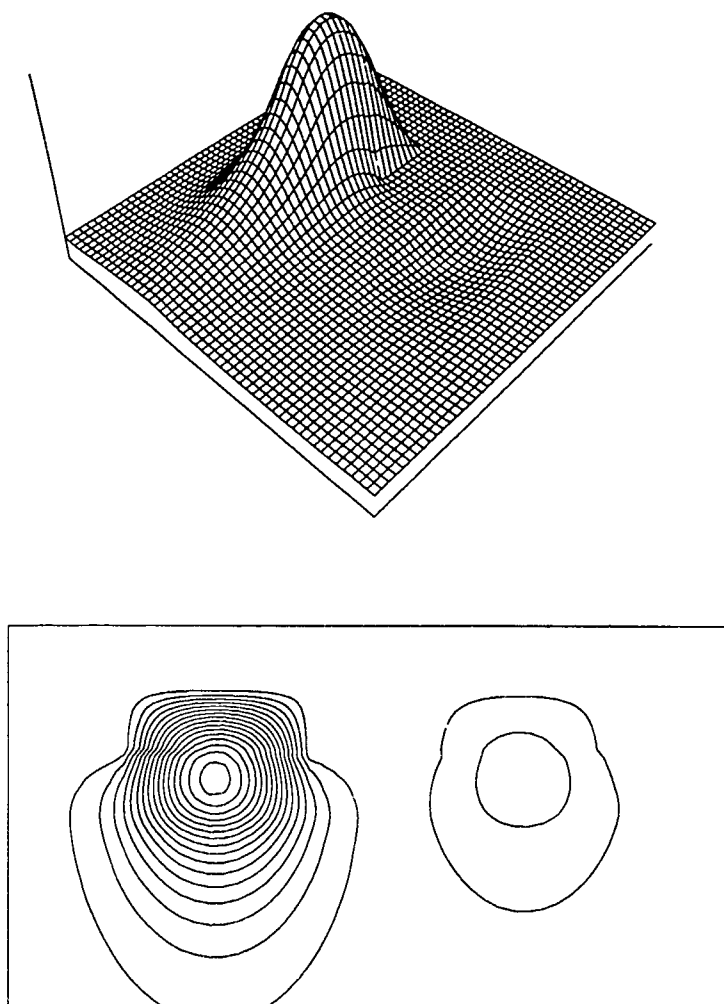


Figure 3. Surface and contour plots of the first higher-order field profile of the dual rib waveguide structure. The parameters are given in the text.

for waveguide 1 and 0.35876 for waveguide 2. By observing the behavior of the field profiles and propagation constants and comparing their characteristics to those of similar yet well-known planar waveguide structures, it may be said that the results are very reasonable. However, it is desirable to compare these results with another method to verify absolute accuracy. The next section will briefly describe the finite element method which, in the future may be used to verify the accuracy of the modified spectral index method.

III. FINITE ELEMENT METHOD

The finite element method (FEM), originating from solid mechanics, is a computer-aided mathematical technique for obtaining approximate numerical solutions to the abstract equations that predict the response of physical systems subjected to external influences [8]. Although the applications of the FEM are quite broad, its implementation may be summarized by six steps:

1. *Discretization of the continuum.* The continuum over which the field is to be determined is broken up into an equivalent system of finite elements. If this is a two-dimensional continuum, it is broken up into a set of logically placed polygons. All polygons must be of the same order, but may be of different size and shape.

2. *Selection of the field variable models.* The field variable over each element is assumed in terms of analytic functions. Usually, a complete set of polynomials are used in which the coefficients are determined by a set of field amplitudes given at the chosen set of nodes on the parameter and/or within the element. The positions of these nodes are determined by the types of approximating functions used.

3. *Derivation of the finite element equations.* Equations are derived for an element relating the nodal field values to the geometric and material properties as well as the nodal values of any applied fields. These equations are expressed as a set of simultaneous linear algebraic equations and thus are expressible in matrix form. A variational method may be used for these relationships.

4. *Assembly of the algebraic equations for the overall discretized continuum.* An overall global matrix equation is constructed from the matrix equations of the individual elements. This construction considers that the nodal points shared by multiple elements have the same value from each element containing that node. Furthermore, all boundary conditions are considered between the elements and the boundaries of the overall structure.

5. *Solution for the nodal field variable vector.* A set of nodal field vectors is given in terms of the eigenvectors of the global matrix equation. Generally, this may be composed of a set of physical solutions as well as a, generally larger, set of non-physical solutions. These are differentiated by use of other field equations not considered in the variational method. The eigenvalues are the mode constants. Each nodal field vector consists of a complete set nodal field values for a given mode. The nodal values may be used along with the approximating functions to determine the approximate field solution over the continuum.

6. *Computation of the element resultants from the nodal field variable amplitudes.* Other types of fields which are related to the one calculated in steps 1 through 5 may be calculated from any equations relating them.

These six steps will be briefly elaborated on as they are applied to the two-dimensional waveguide problem.

STEP 1

Figure 4 shows how a channel waveguide may be broken down into a set of triangular elements. Although only right triangles were used here, there is no restriction placed on the shape or size of a triangle. They must, however share vertexes with neighboring triangles. In Figure 4, a higher density of triangles is used in the region where the solution is expected to be most important. They are arranged so that the geometrical boundaries of the waveguide coincide with edges of some of the triangular elements. This is important because the material properties are assumed to be constant over each element. In a symmetric problem such as the one illustrated, the problem may be reduced in half by imposing boundary conditions along a symmetric wall.

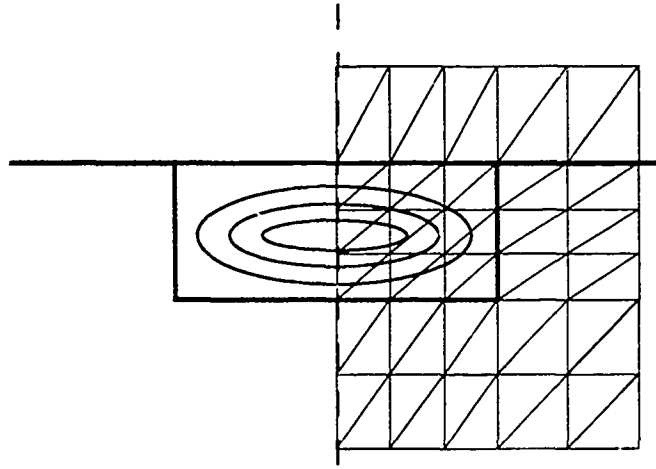


Figure 4. Demonstration of how a channel waveguide may be broken down into a set of triangular elements.

STEP 2

The principle of the finite element method is to approximate the field solution with analytic functions over each element. The problem is then found to be equivalent to finding the extremum of a functional along with the boundary conditions. Any solution may be approximated by relatively low-order polynomials making them a logical choice for this purpose. The polynomial approximations to the field variables of an individual element will now be described.

The field is expressed in terms of a complete polynomial in the transverse plane x and y . As discussed by Dunne [9], in Cartesian polynomial fields, the absence of a preferred direction requires that the polynomial be complete. This property allows the solution to be equal not just at the nodes but also along the entire boundary between two adjacent triangles when one has a different rotation or shape than the other. A complete polynomial of degree N in two dimensions requires $n = \frac{1}{2}(N + 1)(N + 2)$ terms and thus require n parameters for each element. These parameters are given in terms of the field values at n nodes on the boundaries and within the element. Figure 5 shows a placement of n nodes for a 4th order polynomial. If there are no boundary conditions imposed along a triangle edge, the field must

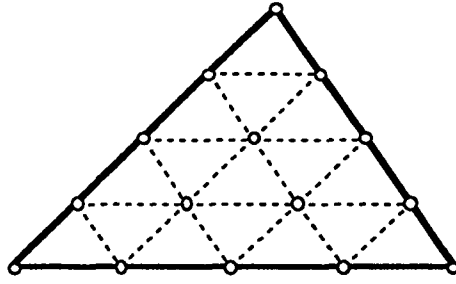


Figure 5. A placement of 15 nodes for a 4th order polynomial. In practice, the nodes could be anywhere within the element, but this arrangement proves to be most convenient.

be continuous across the edge. This requirement is met if the field is specified at the triangle vertices and at $N - 1$ intermediate points along the boundary. Any remaining field values are incorporated within the element with equal spacing by means of triangle area coordinates as was done by Silvester [10].

STEP 3

The next step in the finite element method is to derive the expressions for the coefficients of the polynomials (the nodal values of the field) for each element by utilizing a variational expression. Burk showed [11] that for a general vector \mathbf{H} -field case with \mathbf{H} components in any direction, the problem can be reduced to the determination of the extremum of the functional

$$\omega^2 = \frac{\int (\nabla \times \mathbf{H})^* \cdot \hat{\epsilon}^{-1} (\nabla \times \mathbf{H}) d\Omega}{\int \mathbf{H}^* \cdot \hat{\mu} \mathbf{H} d\Omega}$$

provided the boundary conditions are met. Rahman and Davies showed [2] that by adding a penalty function in the numerator of the functional, spurious solutions may be avoided, and the solution is more well-behaved and accurate. This functional is given by

$$\omega^2 = \frac{\int (\nabla \times H)^* \cdot \hat{\epsilon}^{-1} (\nabla \times H) d\Omega + (\alpha/\epsilon_0) \int (\nabla \cdot H)^* (\nabla \cdot H) d\Omega}{\int H^* \cdot \hat{\mu} H d\Omega}$$

In many finite element applications, the functional are not ratios of functions of the field. Therefore, the problem is reduced to finding the extremum of the functional over each element individually. This leads to the need to solve the eigensystem $Ax = \lambda x$. This functional, however, requires the solution of the eigensystem $Ax = \lambda Bx$.

STEP 4

The matrix equations obtained in step 3 are solutions over each individual element. To obtain the solution of the whole structure, it is necessary to find the solution of all of the elements simultaneously while maintaining that the field values at nodes shared between elements are equivalent. This is done easily by constructing a large diagonal matrix equation from the individual matrix equations and multiplying by a connecting matrix. The result is a sparse matrix.

STEP 5

The eigenvalues and eigenvectors of the matrix equation from step 4 are then found using one of many available computer subroutines. The eigenvectors give the field values at the nodes and the eigenvalues give the propagation constants. Since the number of solutions is equal to the order of the matrix equation and the number of physical solutions may be quite small, the remaining solutions are nonphysical. The problem remains of how to differentiate them. The functional used by Rahman and Davies [2] causes most of the physical solutions to be given in the first solutions, but higher-order solutions may be separated by nonphysical solutions. One way to assure a physical solution is to return to

Maxwell's equations and use $\nabla \cdot E = 0$. The physical solutions will satisfy this equation more accurately.

STEP 6

The method used by Rahman and Davies [1,2,3] determines each of the components of the magnetic field. These may be used to determining the electric field values as well by using Maxwell's equations.

IV. CONCLUSIONS AND FUTURE WORK

The theory as well as a computer program were developed for the determination of the propagation constants and field profiles of multiple waveguide structures using the spectral index method. The results were shown to be very reasonable yielding field profiles suitable for a field propagation program. The finite element method may be used to obtain more accurate solutions at a cost of much greater computer time. However, the accuracy of the spectral index method is believed to be good enough for most purposes. Error due to fabrication is expected to play a greater role. The comparison between the spectral index and the finite element methods has yet to be done to verify this claim.

Results from this work may be incorporated in the design of a broad range of new integrated optical devices such as power dividers, switches, modulators, filters, and multiplexers/demultiplexers. This broad range of applications of multiple rib waveguide structures makes the spectral index method an important tool for the development of new devices.

REFERENCES

- [1] B. M. A. Rahman and J. B. Davies, "Finite Analysis of Optical and Microwave Waveguide Problems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, no. 1, pp. 20-28, 1984.
- [2] B. M. A. Rahman and J. B. Davies, "Penalty Function Improvement of Waveguide Solution by Finite Elements," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 922-928, 1984.
- [3] B. M. A. Rahman and J. B. Davies, "Vector-*H* Finite-Element Solution of GaAs/GaAlAs Rib Waveguides," *Inst. Elec. Eng. Proc.*, vol. 132, pt. J, pp. 349-353, 1985.
- [4] Paul W. A. McIlroy, Michael S. Stern, and Peter C. Kendall, "Spectral Index Method for Polarized Modes in Semiconductor Rib Waveguides," *J. Lightwave Tech.*, vol. 8, no. 1, pp. 113-117, 1990.
- [5] W. K. Burns and A. F. Milton, "Mode Conversion in Planar-Dielectric Separating Waveguides," *IEEE J. Quantum Electron.*, vol. QE-11, pp. 32-39, 1975.
- [6] C. P. Hussell, R. V. Ramaswamy, and R. Srivastava, "Adiabatic Invariance in GRIN Channel Waveguides and its Application to 3dB Cross-Couplers," *Applied Optics*, vol. 29, pp. 4105-4110, 1990.
- [7] Dietrich Marcuse, "Directional Couplers Made of Nonidentical Asymmetric Slabs. Part I: Synchronous Couplers," *J. Lightwave Tech.*, vol. LT-5, no. 1, pp. 113-118, 1987.
- [8] David S. Burnett, *Finite Element Analysis*, Addison-Wesley, Massachusetts, 1987.
- [9] P. C. Dunne, "Complete Polynomial Displacement Fields for Finite Element Method," *The Aeronautical Journal of the Royal Aeronautical Society*, vol. 72, pp. 245-246, 1968.
- [10] P. Silvester, "High-Order Polynomial Triangular Finite Elements for Potential Problems," *Int. J. Engng Sci.*, vol. 7, pp. 849-861, 1969.
- [11] A. D. Berk, "Variational Principles for Electromagnetic Resonators and Waveguides," *IRE Trans. Antennas and Propagation*, vol. AP-4, pp. 104-111, 1956.

MATERIALS FOR INFRARED MODULATOR TECHNOLOGY

Annette M. Marsh

Graduate Student

Abstract

Modulators play a very powerful role in many military applications. A list of possible electrooptic modulator materials was compiled from a literature search. A good modulator material has the characteristic quality of a large electro-optic coefficient or a large Verdet constant. One method of determining the new indices of refraction after an electric field is applied, and the indices are rotated, is by using laser polarimetry. Laser polarimetry at Eglin Air Force Base uses the polarimeter data to form a Mueller matrix. By knowing the crystal structure of the sample and the crystal plane the electric field is applied across, the electrooptic coefficient can easily be calculated. This was demonstrated by measuring the electrooptic coefficient found to be 5.8×10^{-12} m/V, for a cadmium telluride sample. It was determined from a survey of material properties that lithium tantalate show great promise for a modulator operating in the infrared wavelength region.

Introduction

Infrared and optical materials play a large role in the research at Eglin Air Force Base because of their importance in many tactical missile applications. Laser polarimetry got its start with Dr. Dennis Goldstein's research into modulators and modulator materials. Dr. Goldstein laid the groundwork with a thorough investigation of infrared spatial light modulator technology for his dissertation. Infrared laser polarimetry was started in 1988 by Dr. Dennis Goldstein and Mr. David Chenault with extensive research in Mueller matrix polarimetry. Infrared laser polarimetry at Wright Laboratory has made significant advances since its initiation. The work in the early stages by summer graduate students focused on reducing measurement and matrix errors. Progress at that stage was significantly hampered by laser instability, beam wander and inadequate polarizers. Since its initiation, the project advancement has been enhanced by the purchase of a new CO₂ laser and continued computer program improvements. More recently, work has concentrated on retarder error measurements and fast axis alignment, beam wander, detector accuracy, and power output corresponding to particular laser wavelengths.

Infrared laser polarimetry is a process to obtain polarization properties of a material which is transparent to infrared radiation.[1] A polarimeter determines the polarization state of light after it passes through a medium of air or a sample. A polarimeter is designed to operate over different wavelengths depending on the source used. Polarimetry experiments are necessary to document properties of crystals in order to determine if they would be applicable as a suitable electrooptic modulator material.

Electrooptic modulators have a wide range of applications including target simulation, real time radar image processing and optical signal processing systems. Infrared target simulation is an important application for modulators. It is beneficial to

be able to test infrared systems in the laboratory and to do so requires some method of presenting real or synthetic infrared imagery to the systems. Modulators which can serve as extended infrared scene generators are required for this application.[2] In laser radar systems, modulators are used as programmable filters for systems that process range and Doppler information. Optical computing is an area of technology that requires modulators as processing and interconnect elements. Although it is a relatively undeveloped area with many theoretical parameters and practical considerations, the optical transmission, contrast ratio and frame rate capabilities have made noteworthy advances recently. The primary advantages of optical computing are speed and simplicity it offers. This progress has been made in part by improved materials selection, characterization, and film to substrate deposition techniques.

A simple modulator operates by imposing a magnetic or electric field and utilizes basic polarization properties of a crystal. A sample set-up is illustrated in Figure 1. In the modulator, unpolarized light is furnished by an outside source. The light is then polarized by a polarizing film before it reaches the crystal. When the modulator is magnetically controlled, polarized light entering the film will be rotated clockwise or counterclockwise, and the amount of rotation termed the Faraday rotation angle. The direction of rotation is determined by the direction of magnetization. After passing through the film the light enters an analyzer where it will either be blocked or transmitted depending on the direction of polarization. It is through this method that control of pixels to be light, dark or gray is achieved.

The structure of a magneto optic modulator starts with a nonmagnetic substrate. A film is epitaxially deposited in the form of mesas or pixels. These pixels are separated from each other in order that no magnetic field interaction occurs, and they have a thickness of approximately 5-25 microns. The mesas are formed through a series of

photolithographic and etching steps. The magnetization direction of each pixel is individually controlled. The structural path light travels through begins with a polarizer, the pixel (consisting of the substrate and film), and then an analyzer.

The switching of a pixel by electronic or magnetic methods is depicted in Figure 2.[3] There is a patterned two level conductor with a transparent dielectric deposited in between the two conductors. The conductors act as current drivelines and are formed by sputtered gold. The switching process is initiated in the pixel corner, at the intersection of the two conductors. A magnetic field is generated at this point which is large enough to overcome the switching threshold, but is unable to change the state of the mesa. The switching begins by flux reversal at the magnetic domain point. From this point a vertical magnetic domain wall is produced and driven across the pixel. The wall moves across the pixel at a rate of 100 meters per second. The formation of the vertical domain wall requires a few nanoseconds. The entire process takes approximately a few tens of nanoseconds. If the magnetic field is removed before the domain wall has been driven through the film thickness, the mesa will become demagnetized. A saturation field, H_s , is required for complete switching to occur.

Objectives

There is a vast world of available materials that have undocumented properties. This is particularly true for optical materials. It is not uncommon to hear members of the scientific community being quoted expressing their frustration at this deficiency for optic, electrooptic and magneto-optic materials. Materials selection plays a key role in almost every electronic component production and cannot be overlooked. The goal for magneto-optical modulators is a balance of many different parameters. It is generally desired to achieve a maximum Faraday rotation angle with a minimum attenuation, along with temperature independence. One approach to measuring optical properties of

materials is by utilizing laser polarimetry. Laser polarimetry yields information about a material by employing a mathematical construct called a Mueller matrix for crystalline samples. These samples can be in a single crystal or liquid crystal form. The Mueller matrix is a major part of a complete optical description of a medium, and is a sixteen element matrix. From the matrix we are able to determine polarization properties. In addition, the electrooptic coefficients can be found.

My objectives for summer research at Eglin have been taken from a materials science point of view. The overall goal has been to find a good magneto-optical modulator material. Specifically, a material that yields a high Faraday rotation angle, low attenuation and operates in the infrared region of the spectrum. There are considerable advances to be made in the materials area for modulator technology to advance. Not only are there many different possible materials but also film thickness and deposition techniques should not be neglected. The second goal concerns the equations of the indices of refraction. A set of equations for tetragonal crystal structure along individual lattice planes was determined mathematically. In addition, an experiment was performed using the laser polarimeter, testing a cadmium telluride sample. From this it was shown how the electrooptic coefficient could be determined at a wavelength of 10.6 microns.

I have just completed my first semester as a graduate student at the University of Florida in the Materials Science and Engineering department, where I am specializing in electronic materials. My background in electronic materials has dealt with basic materials selection and thin film deposition processing techniques. I am currently pursuing a Masters of Science, specializing in the area of electrooptic materials.

Results

For each crystal sample there exists an index ellipsoid equation. A mathematical method to determine good modulator material starts with an index of refraction equation

which can be derived from the index ellipsoid. When an electric or magnetic field is applied across a crystal, an index change occurs. The new indices of refraction can be found by solving an eigenvalue problem. From these new index equations, it can be seen in each that the primary variable in each is a $n_o^3 r$ term. Therefore to find a material that gives a maximum Faraday rotation, the $n_o^3 r$ terms for materials transmitting at the infrared wavelength are compared. The index equations are sensitive to the crystallographic plane which the electric field is applied across. One problem with this simple comparison is that it neglects the voltage that is required to initiate the switching for each material. It is often the case that it is quite large.

The crystal structure of materials is a influential parameter when considering optical materials. Every material can be classified under one of seven crystal systems. Each crystal system represents a unit cell geometry that is the basic unit or building block that forms a material. The term crystalline material simply means that the unit cell is regular and repeating by stacking these units together. The unit cell describes the axial lengths and angles (see Figure 3). Materials also have a limited number of atom arrangements within a unit cell. The atoms are considered to be points and there are fourteen different arrangements called the Bravais Lattices (refer to Figure 4).

As previously stated, the method of the index of ellipsoid is most commonly used to find the indices of refraction, as in the following case. The energy density of the stored electric field in the anisotropic medium is defined as:

$$U_e = \frac{1}{2} \mathbf{E} \cdot \mathbf{D}$$

From this equation we can write the energy density in terms of D space yielding:

$$\frac{D_x^2}{\epsilon_x} + \frac{D_y^2}{\epsilon_y} + \frac{D_z^2}{\epsilon_z} = 2U_e$$

In this equation, $\epsilon_x, \epsilon_y, \epsilon_z$, are the principal dielectric constants. The next step is to replace the D/\sqrt{U} term by r and define the principal indices of refraction n_x, n_y, n_z , by $n_i^2 = \frac{\epsilon_i}{\epsilon_0}$ ($i=x,y,z$). We now have, with these changes, an equation of a general ellipsoid with major axes parallel to the x, y , and z directions, with respective lengths, $2n_x, 2n_y, 2n_z$; written below:

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1$$

This is the index ellipsoid reduced to the simplest form in the principal coordinate system. The index ellipsoid is used to find the two indices of refraction along with the two associated directions of D relating to the two independent plane waves that propagate along an arbitrary direction S in a crystal.[4] The $\frac{1}{n_x^2}, \frac{1}{n_y^2}$, and $\frac{1}{n_z^2}$ terms are the principal values of the impermeability tensor $\eta_{ij} = \epsilon_0(\epsilon^{(-1)})_{ij}$ where $\epsilon^{(-1)}$ is the inverse of the dielectric tensor. The relation between the field vectors E and D can be written:

$$E = \frac{1}{\epsilon_0} \eta D$$

According to quantum theory, the optical dielectric impermeability tensor is dependent on the charge distribution in the crystal. When an electric field is applied, a redistribution of the bond charges and possibly a deformation of the ion lattice will occur. This results in a change in the optical impermeability tensor. This is the electrooptic effect. Electrooptic coefficients are defined in terms of the impermeability tensor and the electric field. Now the index ellipsoid of a crystal with an electric field applied is defined as:

$$\eta_{ij}(E)x_i x_j = 1$$

If the electric field is removed the index ellipsoid reduces back to its simplest form.

Because the dielectric tensor and the impermeability tensor are symmetric, the electrooptic coefficients can be written in a contracted notation. This is usually in matrix form for specific crystal symmetry classes, such as tetragonal for a $\bar{4}2m$ point group:

$$r_{ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{63} \end{pmatrix}$$

In this form, it is clearly illustrated that r_{41} , r_{52} , and r_{63} are the only non-zero elements.

The other elements vanish due to symmetry in the crystal. Now the index ellipsoid equation with the presence of an electric field can be written in general form using the contracted indices:

$$\left(\frac{1}{n_x^2} + r_{1k} E_k \right) x^2 + \left(\frac{1}{n_y^2} + r_{2k} E_k \right) y^2 + \left(\frac{1}{n_z^2} + r_{3k} E_k \right) z^2 + 2yzr_{4k} E_k + 2zxr_{5k} E_k + 2xyr_{6k} E_k = 1$$

where E_k ($k=1, 2, 3$) is a component of the electric field and summation of repeated indices k is assumed. For the tetragonal $\bar{4}2m$ material, the index ellipsoid equation now becomes:

$$\frac{x^2}{n_o^2} + \frac{y^2}{n_o^2} + \frac{z^2}{n_e^2} + 2r_{41} E_x yz + 2r_{41} E_y xz + 2r_{63} E_z xy = 1$$

The index ellipsoid for an electric field applied along different crystal planes can now be determined. The eigenvalue problem is then formulated as follows:

$$\begin{pmatrix} \frac{1}{n_o^2} & 0 & r_{41} \frac{E}{\sqrt{2}} \\ 0 & \frac{1}{n_o^2} & r_{41} \frac{E}{\sqrt{2}} \\ r_{41} \frac{E}{\sqrt{2}} & r_{41} \frac{E}{\sqrt{2}} & \frac{1}{n_e^2} \end{pmatrix} \vec{V} = \lambda \vec{V}$$

The secular equation is determined by subtracting the $\frac{1}{n^2}$ term from the diagonal

elements:

$$\begin{pmatrix} \frac{1}{n_o^2} - \frac{1}{n^2} & 0 & r_{41} \frac{E}{\sqrt{2}} \\ 0 & \frac{1}{n_o^2} - \frac{1}{n^2} & r_{41} \frac{E}{\sqrt{2}} \\ r_{41} \frac{E}{\sqrt{2}} & r_{41} \frac{E}{\sqrt{2}} & \frac{1}{n_e^2} - \frac{1}{n^2} \end{pmatrix}$$

The roots of this matrix are the new principal indices after an electric field has been applied. These results for tetragonal crystal systems were calculated and are presented in Table 1.

The next task was to research and compile information on known electrooptic materials. An extensive literature search was performed to gather this information consisting of basic material properties. The results were assembled in tabular form and are listed in Table 2[5] and Table 3[6]. It should be stressed that this is a list of commonly termed "electrooptic" materials, there still remain other materials that are not listed.

There is a significant lack of electrooptic coefficients values available, as illustrated. The electrooptic coefficient can be determined experimentally using laser polarimetry and the Mueller matrix. This information would be extremely beneficial in

determining the success of possible modulator materials. The determination of the electrooptic coefficient was performed using laser polarimetry for a cadmium telluride sample. The electric field was applied across the (110) crystal plane. Voltage readings on the sample were taken ranging from 100 to 3000 Volts. From this data, a Mueller matrix was derived. The fast axis was set at zero degrees and the laser wavelength at 10.6 microns. The m_{44} element of the matrix was used in calculating the electrooptic coefficient. The graph was plotted using the following equation:

$$M_{44} = \cos \frac{2\pi}{\lambda} n^3 r_{41} V \frac{L}{d}$$

Because we know the electrooptic coefficient is equal to $n^3 r_{41}$, plotting the inverse cosine of the M_{44} term versus the $2\pi \frac{L}{\lambda} V \frac{L}{d}$ yields a slope equal to the $n^3 r$ value from which the EO coefficient can easily be calculated (see Figure 5). In Figure 5, the theoretical values are plotted as the straight line and the experimental values as the curved line. This graph illustrates the accuracy of the laser polarimetry method of determining EO coefficients. The experimental electrooptic coefficient value for CdTe is 5.8×10^{-12} m/V, and the theoretical value calculated from the half wave voltage data is 5.7×10^{-12} . This illustrates that by knowing the crystal symmetry of a material and the crystal plane the electric field is being applied across, laser polarimetry can be used to obtain optical properties such as the electrooptic coefficient, making it a very effective tool.

Conclusions

In conclusion, infrared laser polarimetry is a very powerful tool that can be used to document desperately needed optical material properties. This technology is essential for growth in the field of electrooptic and magneto-optic modulators. According to available

data, two promising materials for modulator applications appear to be lithium niobate and lithium tantalate for a wavelength of 3.39 microns. Further investigation and experimentation is needed on many different materials in order to document basic properties such as the electrooptic coefficient, in order for future advancement. In addition, determining the index ellipsoid equations for other crystal structures over a wide range of crystal planes would also be greatly beneficial in finding other modulator materials that would yield a maximum Faraday rotation angle. These ellipsoid equations would allow the new indices of refraction to be determined. The future of optical processing systems and radar imaging systems is promising but will require some basic characterization of materials in order to make significant advances.

Acknowledgments

I would like to thank Research and Development Laboratories for sponsoring this program and for the time and interest they took in the participants of the program. I would also like to thank the members of the special projects laboratory, Howard McCormick, Michael VanTassel, Voncile Houston, David Onuffer, and Linda Lau for all their assistance. My deepest appreciation goes to my focal point, Dr. Dennis Goldstein, for the opportunity to work in opto-electronics and for all the knowledge and assistance throughout the summer. Finally, a special thanks to all my fellow students for their help and a great summer at Eglin, Lisa Collins, Chad Houghton, Danielle Walker, and especially Randy Gove for his additional guidance with the laser.

Table 1: Electrooptical Properties of Tetragonal $\bar{4}2m$ Crystals

E perpendicular to $\langle 001 \rangle$ plane $E_x = E_y = 0, E_z = E$

Index Ellipsoid:
$$\frac{x^2 + y^2 + z^2}{n_o^2} + 2r_{63}Exy = 1$$

$$n_x' = n_o + \frac{1}{2}n_o^3r_{63}E$$

$$n_y' = n_o - \frac{1}{2}n_o^3r_{63}E$$

$$n_z' = n_e$$

E perpendicular to $\langle 110 \rangle$ plane $E_x = E_y = E/\sqrt{2}, E_z = 0$

Index Ellipsoid:
$$\frac{x^2 + y^2 + z^2}{n_o^2} + \sqrt{2}r_{41}E(yz + zx) = 1$$

$$n_x' = n_o + \frac{1}{2}n_o^3r_{41}E$$

$$n_y' = n_o - \frac{1}{2}n_o^3r_{41}E$$

$$n_z' = n_e$$

E perpendicular to $\langle 111 \rangle$ plane $E_x = E_y = E_z = E/\sqrt{3}$

Index Ellipsoid:
$$\frac{x^2 + y^2 + z^2}{n_o^2} + \frac{2}{\sqrt{3}}r_{41}E(yz + zx) + \frac{2}{\sqrt{3}}r_{63}Exy = 1$$

$$n_x' = n_o + \frac{1}{4\sqrt{3}}n_o^3E(2r_{63}^2 - 8r_{41}^2)^{1/2}$$

$$n_y' = n_o - \frac{1}{4\sqrt{3}}n_o^3E(2r_{63}^2 - 8r_{41}^2)^{1/2}$$

$$n_z' = n_e + \frac{1}{2}n_o^3\frac{r_{63}}{\sqrt{3}}E$$

TABLE 2: INFRARED TRANSMITTING MATERIALS

Material	Trans- mission range	Index of refraction at wavelength of	EO coeff ($\times 10^{-12}$ m/V)
GaAs	2.9 1.15 3.39 10.6	n=3.60 n=3.43 n=3.30 n=3.30	(T) r41=1.10 (T) r41=1.43 (T) r41=1.24 (T) r41=1.51
LiNbO ₃	0.633 3.39	n _o =2.266 n _e =2.200	(T) r13=2.00 (S) r13=6.0 (T) r22=6.80 (S) r22=3.4 (T) r33=30.9 (S) r33=30.8 (T) r51=32.6 (S) r51=28.0 (S) r22=3.1
LiTaO ₃	0.633 3.39	n _o =2.176 n _e =2.180	(T) r13=22.0 (S) r13=7.0 (S) r22=1.0 (S) r33=30.3 (S) r51=20.0 (S) r33=27.0 (S) r22=0.3 (S) r13=45 (S) r51=15.0
ZnS	0.633	n _o =2.347 n _e =2.360	(S) r13=0.9 (S) r33=1.8
ZnSe	0.548 0.633 10.6	n=2.66 n=2.60 n=2.39	(T) r41=2.0 (S) r41=2.0 (T) r41=2.2
ZnTe	0.539 0.616 0.633 0.690 3.41 10.6	n=3.06 n=3.01 n=2.99 n=2.93 n=2.70 n=2.70	(T) r41=4.51 (T) r41=4.27 (T) r41=4.04 (S) r41=4.3 (T) r41=3.97 (T) r41=4.20 (T) r41=3.90
CdS	0.589 0.633 1.15 3.39 10.6	n _o =2.501 n _e =2.519 n _o =2.460 n _e =2.477 n _o =2.276 n _e =2.292 n _o =2.226 n _e =2.239 n _o =2.336 n _e =2.336	(T) r51=3.7 (T) r51=1.6 (T) r31=3.10 r33=3.20 r51=2.0 (T) r13=3.50 r33=2.90 r51=2.0 (T) r13=2.45 r33=2.75 r51=1.7
CdSe	3.39	n _o =2.452 n _e =2.471	(S) r13=1.8 (T) r33=4.3
CdTe	1.0 3.39 10.6 23.35 27.95	n=2.84 n=2.60 n=2.58 n=2.53	(T) r41=4.5 (T) r41=6.8 (T) r41=6.8 (T) r41=5.47 (T) r41=5.04

TABLE 3: MATERIALS LIST

Material	crystalline system	hardness (knoop g/mm ²)	density (g/cc)	water solubility (g/100g water)	transmission range	index of refraction at wavelength of
Al ₂ O ₃	hexagonal	2200	3.97	insol.	0.14-6.5	1.768
BaF ₂	cubic	82	4.89	0.12	0.25-15.0	1.475 at 0.546um
CaF ₂	fc cubic	158	3.18	0.0016	0.13-12.0	1.424 at 2.0um
CsBr	cubic	19.5	4.44	124.3	0.3-55.0	1.671 at 2.0um
CsCl	cubic		3.988	162.2	0.19-30	1.822 at 0.2um
CsI			4.51	44	0.25-80.0	1.747 at 2.0um
Ge	cubic	780	5.323	insol.	1.8-23.0	4.116 at 2.0um
KRS-5		40	7.371	0.05	0.6-40.0	2.395 at 2.0um
LaF ₃			5.94	insol.	0.4-11.0	1.82 at 0.58um
LiBr	cubic		3.464	145		1.784
LiCl	cubic		2.068	63.7		1.662
LiF	cubic	102	2.635	0.27	0.12-9.0	1.379 at 2.0um
MgF ₂	tetragonal	415	3.15	0.0076	0.11-7.5	1.378 at 0.58um
KBr	cubic	7	2.75	53.5	0.25-40.0	1.538 at 2.0um
KCl	cubic	9	1.984	34.7	0.21-30.0	1.475 at 2.0um
KI	cubic	5	3.12	127.5	0.25-45.0	1.631 at 2.0um
RbBr	cubic		3.35	98		1.57 at 0.5um
RbCl	cubic		2.80	77		1.5 at 0.5um
RbI	cubic		3.55	152		1.67 at 0.5um
Si	cubic	1150	2.328	insol.	1.2-15.0	3.458 at 2.0um
AgBr		7	6.473	0.0084	0.5-35.0	2.253 at 0.6um
AgCl	cubic	9.5	5.56	0.0089	0.4-28.0	2.006 at 2.0um
NaBr	cubic		3.203	116	0.22-23	1.64 at 0.6um
NaCl	fc cubic	18	2.165	35.7	0.21-26.0	1.527 at 2.0um

materials	crystal system	hardness (knoop g/mm ²)	density (g/cc)	water solubility (g/100g water)	transmission range	index of refraction at wavelength of
NaF	cubic	60	2.588	4.22	0.13-12	1.317 at 2.0um
NaI	cubic		3.667	184	0.25-25	1.775 at 0.589um
NaN ₃	trigonal	19.2	2.261	92.1	0.35-3.0	1.586 at 0.578um
SrF ₂	cubic	130	4.24	0.011	0.12-11	1.439 at 0.546um
SrTiO ₃	cubic	595	5.175	insol.	0.39-6.8	2.409
TiBr		12	7.557	0.05	0.42-48.0	2.350 at 0.75um
TiCl	hexagonal	13	7.004	0.29	0.40-30	2.198 at 0.75um
TiO ₂	tetragonal	700	4.26	insol.	0.43-6.2	2.613, 2.909
					Point Symm.	see table 2 for index of refraction and EO coeff values
CdS	hexagonal		4.82		6mm	
CdSe	hexagonal		5.81		6mm	
CdTe	cubic	56	6.20	insol.	$\bar{4}3m$	
GaAs	cubic	721	5.30	insol.	$\bar{4}3m$	
LiNbO ₃	trigonal				3m	
LiTaO ₃	trigonal				3m	
ZnS	hexagonal	178	4.09		6mm	
ZnSe	cubic	137	5.42		$\bar{4}3m$	
ZnTe	cubic	82	6.0	insol.	$\bar{4}3m$	

FIGURE 1: Operation of modulator pixels

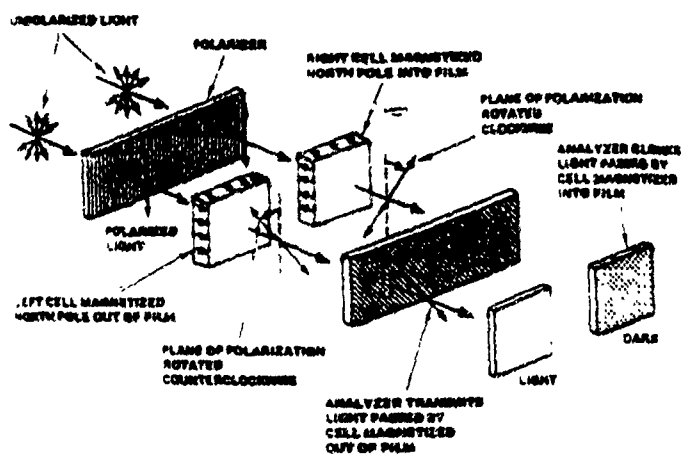


FIGURE 2. Magnetic switching process across a mesa

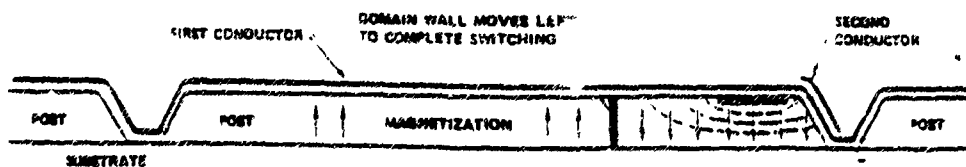


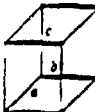



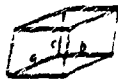


FIGURE 3: The seven crystal systems

System	Axial Lengths and Angles	Unit Cell Geometry
Cubic	$a = b = c, \alpha = \beta = \gamma = 90^\circ$	
Tetragonal	$a = b \neq c, \alpha = \beta = \gamma = 90^\circ$	
Orthorhombic	$a \neq b \neq c, \alpha = \beta = \gamma = 90^\circ$	
Rhombohedral	$a = b = c, \alpha = \beta = \gamma \neq 90^\circ$	
Hexagonal	$a = b \neq c, \alpha = \beta = 90^\circ \neq \gamma = 120^\circ$	
Monoclinic	$a \neq b \neq c, \alpha = \gamma = 90^\circ \neq \beta$	
Triclinic	$a \neq b \neq c, \alpha \neq \beta \neq \gamma \neq 90^\circ$	

The lattice parameters a , b , and c are unit cell edge lengths. The lattice parameters α , β , and γ are angles between adjacent unit cell axes where α is the angle viewed along the a axis (i.e., the angle between the b and c axes). The inequality sign (\neq) means the equality is not required. Accidental equality occasionally occurs in some structures.

FIGURE 4: The fourteen Bravais Lattices

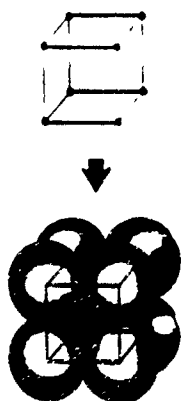
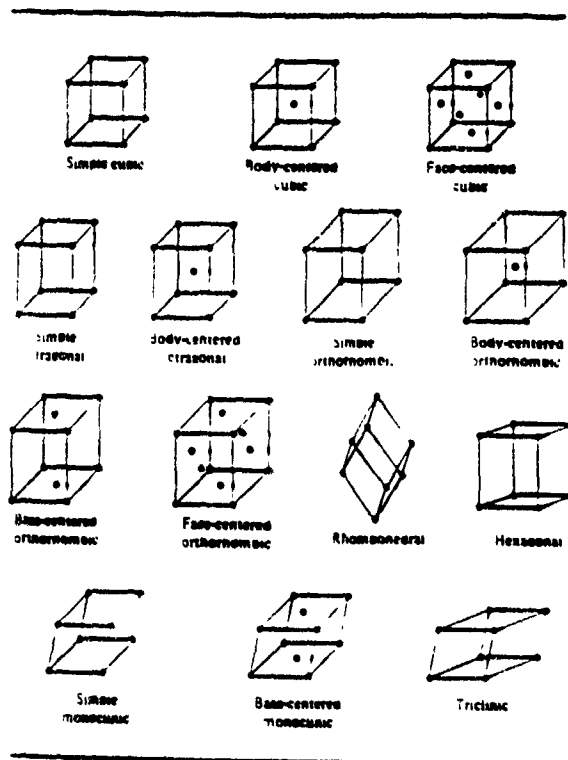
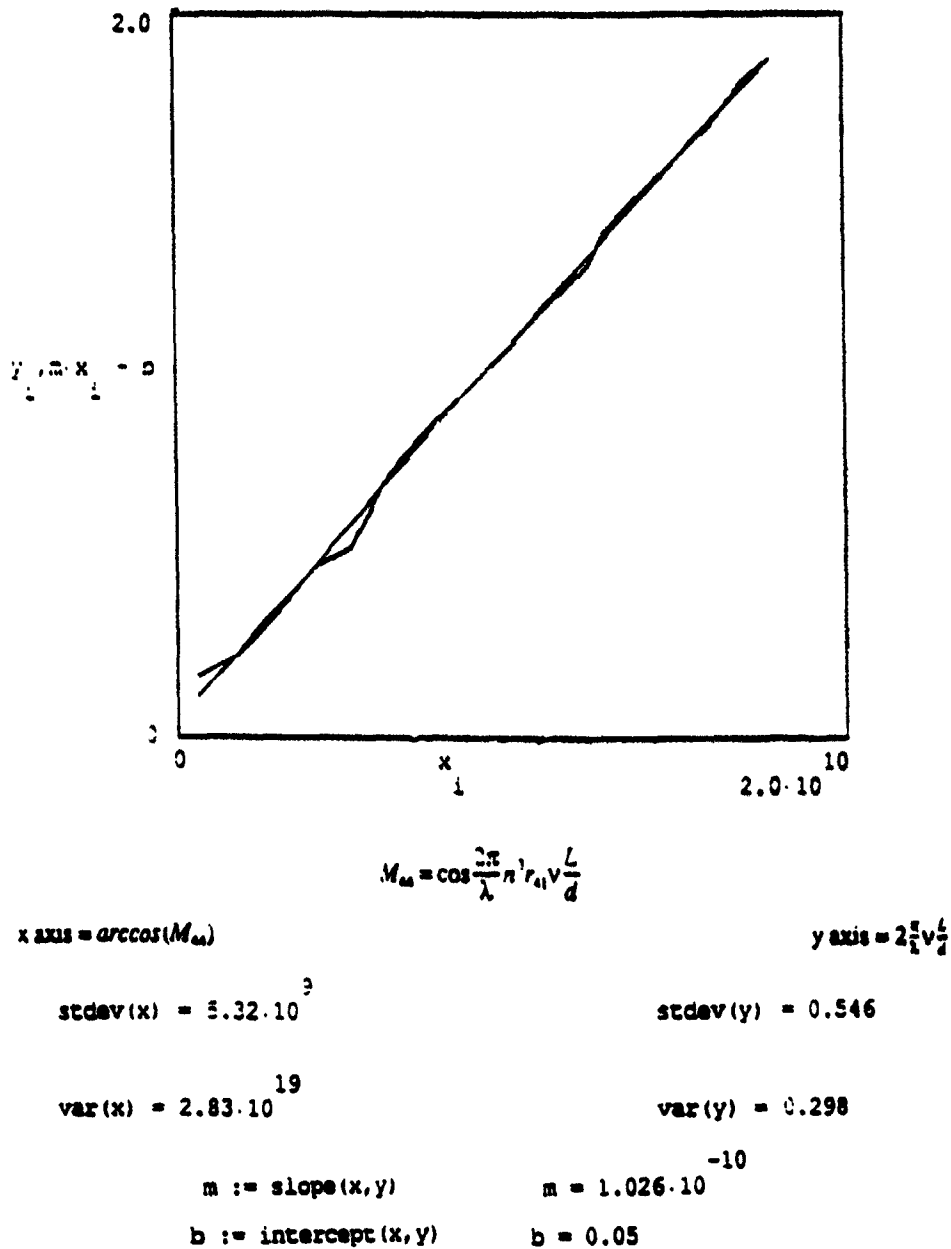


FIGURE 5: Graph of Cadmium Telluride sample
with electric field across (110) plane



Endnotes

1. Gove, Randy., "Infrared Laser Polarimetry", Final Report for the 1990 USAF-UES Graduate Student Research Program, 22 August 1990.
2. Goldstein, Dennis H., "Polarization Modulation in Infrared Electrooptical Materials", Dissertation, University of Alabama in Huntsville, 1990.
3. W.E. Ross, K.M. Snapp, and R.H. Anderson, "Fundemental Characteristics of the Litton iron garnet magneto-optic spatial light modulator", Litton Systems Inc., Data Systems Division.
4. Yariv, Amnon and Yeh, Pochi, "Optical Waves in Crystals", John Wiley and Sons, Inc., New York, 1984.
5. William L. Wolfe and George J. Zissis, The Infrared Handbook, Environmental Research Institute of Michigan, 1989.
6. CRC Handbook of Chemistry and Physics, 70th Edition, The Chemical Rubber Company Publishing, 1989.

MATERIAL CHARACTERIZATION AND EVALUATION FOR
PENETRATOR APPLICATIONS

Graduate Student Jessica L. Mayes

As an AFOSR summer research graduate student, I was involved in several aspects of engineering research. In addition to assisting with various projects in the lab, my main focus was dynamic penetration mechanics relating to the advancement of weapons, specifically warheads. With the aid of good technical equipment and proficient research engineers, significant advancements were made that can be continued upon return to the University of Kentucky's research environment.

In order to make rapid advancements in the realm of warheads, there is a distinct need for development of a three dimensional model to determine the characteristics of materials deformed under high strain rate conditions. Being able to accurately predict the final state of a material eliminates much of the wasted time and funding involved with trial and error.

In researching this problem, it is extremely important to obtain accurate physical data to correlate with dependable hydrocodes. Precisely controlled initial material characteristics are important in order to eliminate comparison errors. This was studied by my co-researcher and fellow graduate student, Steve Hatfield. Equally important is carefully controlled and documented testing conditions to eliminate any unnecessary variables. Once this is accomplished, the samples are returned to the laboratory for post-impact analysis where my concentration for summer research became documenting the microstructure of Rod on Rod Impact specimens and evaluating the findings.

First in post-impact analysis, testing conditions must be considered. The tests involved in this evaluation were the Classic Taylor Test and the Symmetric Rod Impact Test (See figure 1). The Classic Taylor Test

involves a specimen being impacted against a rigid wall. The Symmetric Rod Test involves a one inch impactor being impelled into a six inch receptor of the same material and cross-sectional area. Taylor's one dimensional analysis utilized post-impact geometry to estimate the dynamic strength. A two dimensional analysis using continuum codes was developed to allow comparison of experimental and predicted results. Both methods continue in research today.

The Rod on Rod Impact test brings several advantages. Most importantly, it eliminates a mismatch impedance between the impactor and the anvil, allowing the computer codes to continue with the same material characteristics throughout impact. The symmetric specimens allow a smooth momentum transfer and symmetric deformation of the rods. With this test, anvil compliance problems are overcome, preventing errors due to the plastic deformation of the anvil. Although they have been shown nominal, any frictional effects are also eliminated in Rod on Rod testing.

I approached this problem desiring microanalysis because of its versatile benefit possibilities. By looking at grain trends in the material, one can be furnished a suggestion of the thermomechanical history of the material. Temperature and pressure gradients

across the surface may also become apparent.

Microanalysis can also offer suggestions for wave propagation effects as well as emphasize any physical abnormalities or points of interest.

In order to isolate and study the areas of interest, I sectioned the specimens as shown in figure 2. ROR 2 was cut in half along the axis so that the center plane from impactor through receptor would be exposed. ROR 6 arrived in four separate sections due to the placement of Manganan gages in the receptor to monitor the pressure trends. These gages were placed at 2mm, 7mm, and 3in (one half the length of the receptor). In both specimens, the bottom half was mounted, polished and etched for examination. The top halves were polished using a different method in order to verify that any unique characteristics observed were not caused by the preparation process.

The results from the Manganan gages in ROR 6 can be observed in figure 3. These devices are unfortunately subject to the physical restrictions of the nature of the apparatus. Manganan gages are designed for compression testing only and do not record the significant tensile wave created immediately following impact. Also, these gages typically fail shortly into the test due to the extreme forces of impact. Figure 4

represents the results predicted by the hydrocode calculations. These compression forces are an accurate prediction of the observed values, as represented by points on the graph. In addition to the extension of available data points in computer calculations (beyond where the actual gages fail), the codes produce expected tensile force data. This is extremely important to consider in analyzing the microstructure.

To document and more closely study the microstructure of the deformed rod, I constructed a montage of both ROR 2 and ROR 6 at 50X magnification. This project was well received by my focal point (Joel House) and retained for the laboratory's use and an important integral part of my presentation. For purposes of this report, a written description will suffice.

In ROR 6, the 2mm section of the receptor welded to the impactor. Complete welding was observed on one side with a crack continuing through the remainder of the impact face. This welding provided for some energy absorption. The additional free surfaces created to accommodate the Manganan gages provided a quite appropriate method of internal stress relief. These discrepancies may account for the reduced amount of observed damage (fewer voids and severely deformed

grains) along the axis in ROR 6 versus that seen in ROR 2.

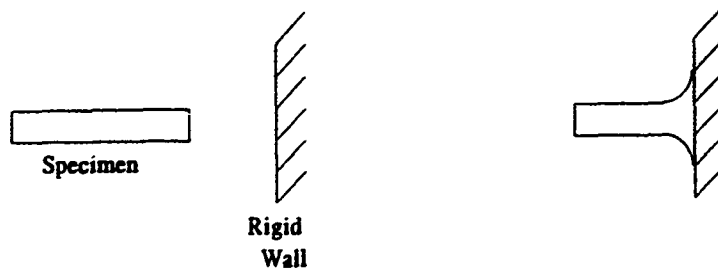
Moving to ROR 2, a simplified overview of the microstructure can be seen in figure 5. The disc-shaped grains parallel to the impact face (three dimensionally, these are pancake-shaped, compressed grains) result from the high compressive axial loads created at impact. The force in this area allowed the grains to compensate in only the horizontal plane, creating the plastic region. Moving back from the impact edge, lower levels of deformation remain primarily because the pressure wave has been sufficiently damped by the time it reaches this area.

Of primary microanalytical interest was the concentration of voids in the central axis region. The high compressive load upon impact is immediately reflected as a strong tensile wave causing void nucleation. This phenomena is observed primarily at the weakest points in the material (usually the grain boundaries). If the amplitude of the tensile wave is sufficient, nucleated voids may link together and become fracture surfaces.

For explanations, spallation fracture was first considered as the primary source of these voids.

However, after careful review of technical literature, it was revealed that as a result of excessive velocity, tensile voids could be created early due to the focusing of radial release waves. Voids created by each of these methods look virtually identical. Also, for comparison, another montage was observed (also at 50X magnification). This showed voids formed in a Classic Taylor-Anvil Test shot at 190m/s. The research group is still in the process of analyzing these voids, as well as the overall microstructure, in both the Symmetric Rod Impact Test and the Classic Taylor Test in hopes of obtaining more definite answers and producing more reliable multi-dimensional models.

Classic Taylor Test



Symmetric Rod Impact Test

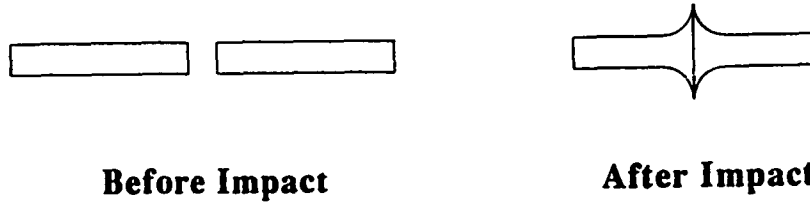
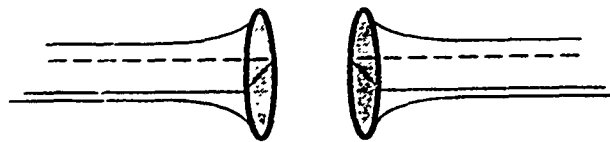
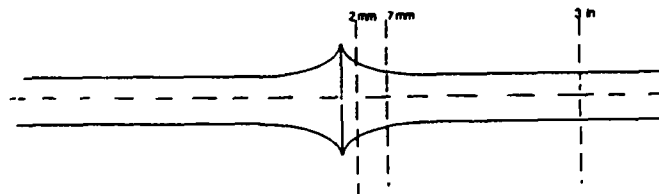


Figure 1 : The Classic Taylor-Anvil Test and The Symmetric Rod on Rod Impact Test



ROR 2



ROR 6

Figure 2 : Isolation and Magnification of Areas of Interest

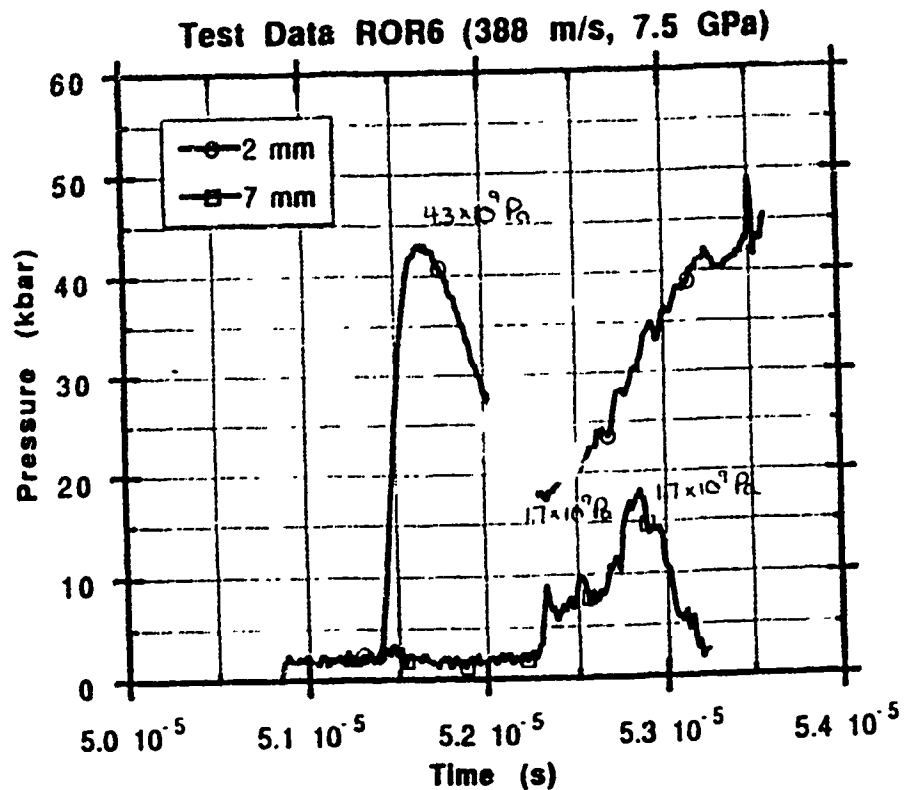


Figure 3 : Actual test data from Manganan gages placed in ROR 6

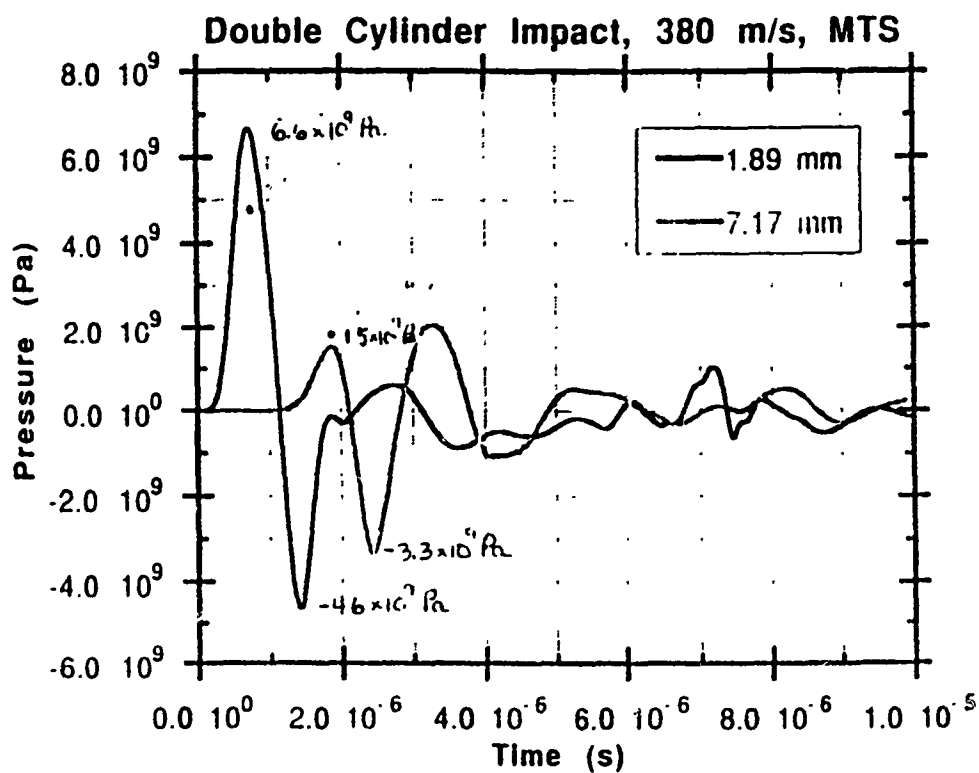


Figure 4 : Predicted pressure values by MTS model

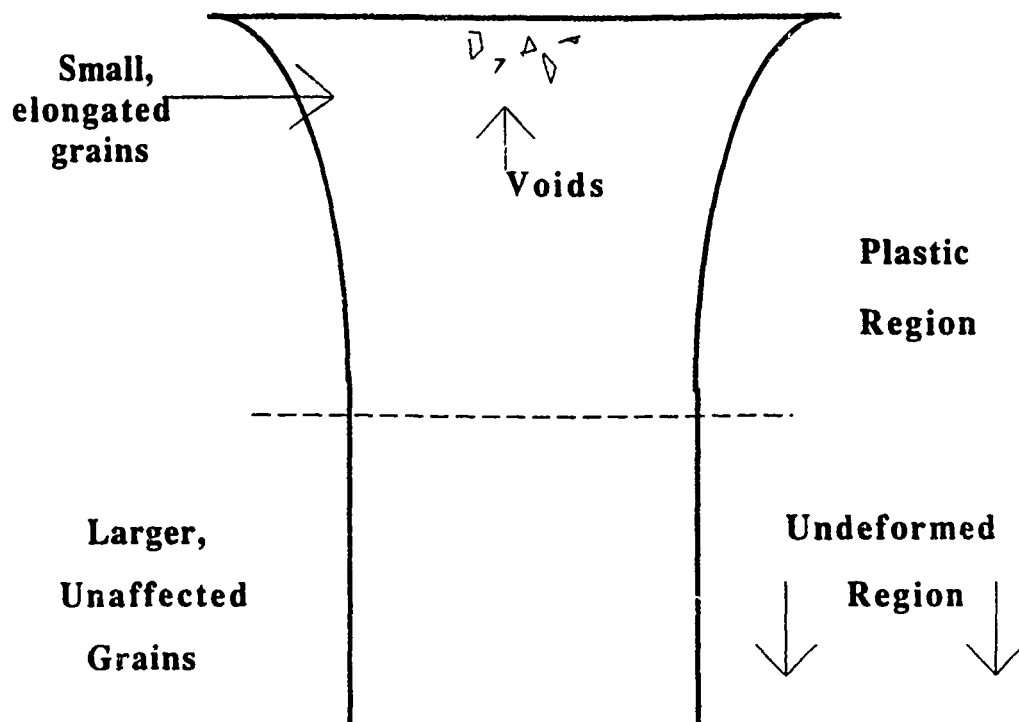


Figure 5 : Overview of areas of interest

FULL REYNOLDS-AVERAGED NAVIER-STOKES SIMULATION
OF TURBULENT TRANSONIC FLOW PAST
A WRAP-AROUND FIN PROJECTILE

Juan B. Ordonez
RDL Research Associate
University of Florida

ABSTRACT

An algebraic grid network has been generated for the numerical solution of the turbulent transonic flow past a wrap-around fin projectile. This network is comprised of four 103x31x55 blocks, with the proper clustering needed to resolve viscous effects at all solid surfaces. The blocks are being solved using a full Reynolds-averaged Navier-Stokes code with the Baldwin and Lomax turbulence model, at a Mach number of 0.95 and an angle of attack of zero degrees. This investigation has not been completed due to errors encountered during preliminary runs. However, intermediate results suggest that the code is robust enough to model the flow in the high curvature regions near the fins.

INTRODUCTION

"Thin-layer" Navier-Stokes codes with the Baldwin-Lomax turbulence model have been proven very effective in the numerical simulation of high speed turbulent flows. These codes, however, are limited by the fact that they can use only one surface in computational space to compute the required

value of the eddy viscosity. This means that the boundary-fitted curvilinear coordinate system to be used in transforming the physical domain must be defined in such a way that all solid surfaces are mapped into a single computational plane. For relatively simple three-dimensional configurations, this is usually accomplished by dividing the entire flow domain into several subdomains, or blocks, so that each block isolates a portion of a solid surface in the configuration of interest. The converged solution to the entire flow field is then obtained by solving these blocks independently while periodically updating the boundary conditions at the corresponding block interfaces. For complex configurations such as a finned missile, however, the restriction to one no-slip computational plane per block is sometimes impossible to implement. At best, this restriction often results in a deterioration of the quality of the grid network in key areas of the flow regime.

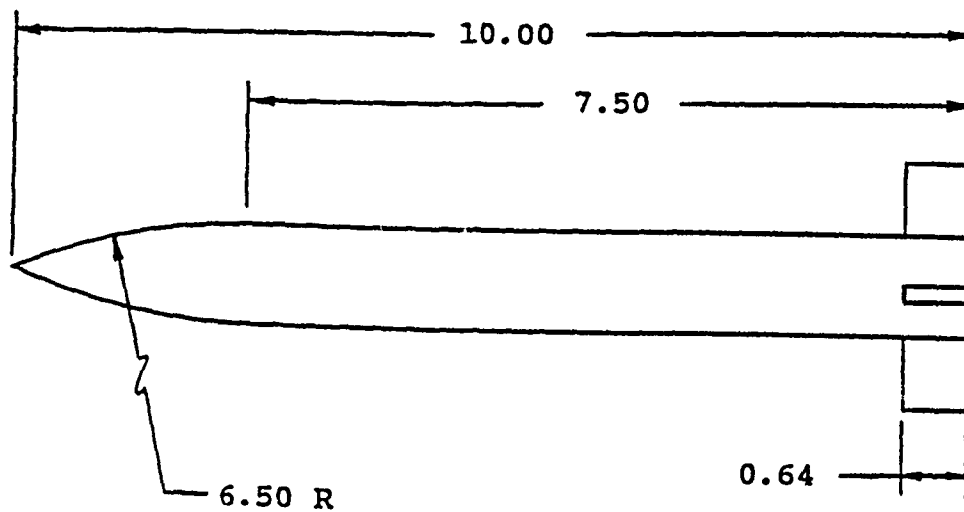
Because of the difficulties and limitations associated with "thin-layer" numerical solutions of complex turbulent flows, it is often necessary to use a Navier-Stokes code having the full set of viscous terms. Full Navier-Stokes codes invariably require more CPU time than the simpler "thin-layer" approximations. However, the inclusion of the extra viscous terms in the governing equations gives the code the distinct ability to accommodate no-slip surfaces in any of the six bounding computational planes of a block. This not only results in a significant simplification of the required block transformations, but it also ensures that the contributions from relevant surfaces affecting the eddy viscosity at any given grid location are all taken into account.

In the present study, the full Reynolds-averaged Navier-Stokes (FRANS)

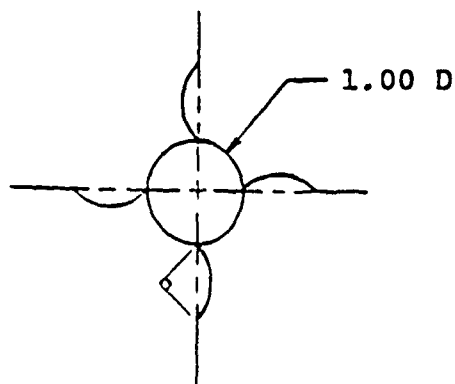
code recently developed at the computational fluid dynamics (CFD) section of the Air Force Armament Laboratory is being used to compute the turbulent transonic flow past a "wrap-around" fin configuration. This code is an extension of the implicit Euler solver of program EAGLE¹, with the full set of viscous terms added explicitly. In addition to having all of the powerful multi-block features of the EAGLE solver, the FRANS code also has the ability to implement the Baldwin-Lomax turbulence model separately in each of the curvilinear directions. The purpose of the current investigation is two-fold: (1) to further test the accuracy, efficiency, and robustness of the FRANS code, and (2) also to study the aerodynamic characteristics of a projectile with curved fins. The wrap-around fin configuration was designed so that each fin wraps around one quarter of the circumference of the projectile. This feature significantly reduces the storage space requirements for the projectile, and also allows for the launching of the projectile directly from a tube. These practical advantages must be weighed against the aerodynamic performance of the projectile. Specifically, an effort is being made to investigate the considerable rolling moment induced by the fin curvature, and the associated roll reversal versus Mach number effect observed during wind tunnel and firing range tests^{2,3}.

GRID TOPOLOGY

As shown in Figure 1, the geometrical configuration considered in this study is a ten-caliber wrap-around fin (WAF) projectile with a 2.5 caliber nose section, where one caliber is defined equal to the diameter of the



a. Side View



b. Rear View

Figure 1. Wrap-around Fin Projectile

projectile body. The support sting, which is not shown in the figure, is twenty calibers long and it is modeled as an extension of the projectile body. As mentioned before, each fin represents one quarter of the circumference of the body. When fully extended, each fin is positioned so that a plane passing through both the fin tip and the fin root is perpendicular to the projectile body at the fin root.

Experience has shown that one of the most important factors contributing to the success of the numerical solution of turbulent flows is the quality of the grid network generated for the problem of interest. A grid network is generally expected to exhibit the following characteristics: (1) a smooth variation of cell volumes and grid line slope; (2) a nearly orthogonal intersection of the grid lines from different curvilinear directions; and (3) an appropriate distribution of points that adapts to steep solution gradients. For a complex three-dimensional configuration, a suitable single grid network which has all of the properties listed above is very difficult to generate. A better approach is to subdivide the flow field into several blocks so that each is generated according to the requirements of its particular subregion. This multi-block approach not only improves the quality of the grid mesh in each region of the flow field, but it also reduces storage and CPU time requirements since only one block is solved at a time.

In the present study, the entire physical domain is divided into four blocks, each spanning the region included between two consecutive fins. This is illustrated in Figure 2. Note that because of the nature of the configuration, blocks 2, 3, and 4 are identical to block 1 except, of course, for their corresponding orientations. Figure 3 shows the physical

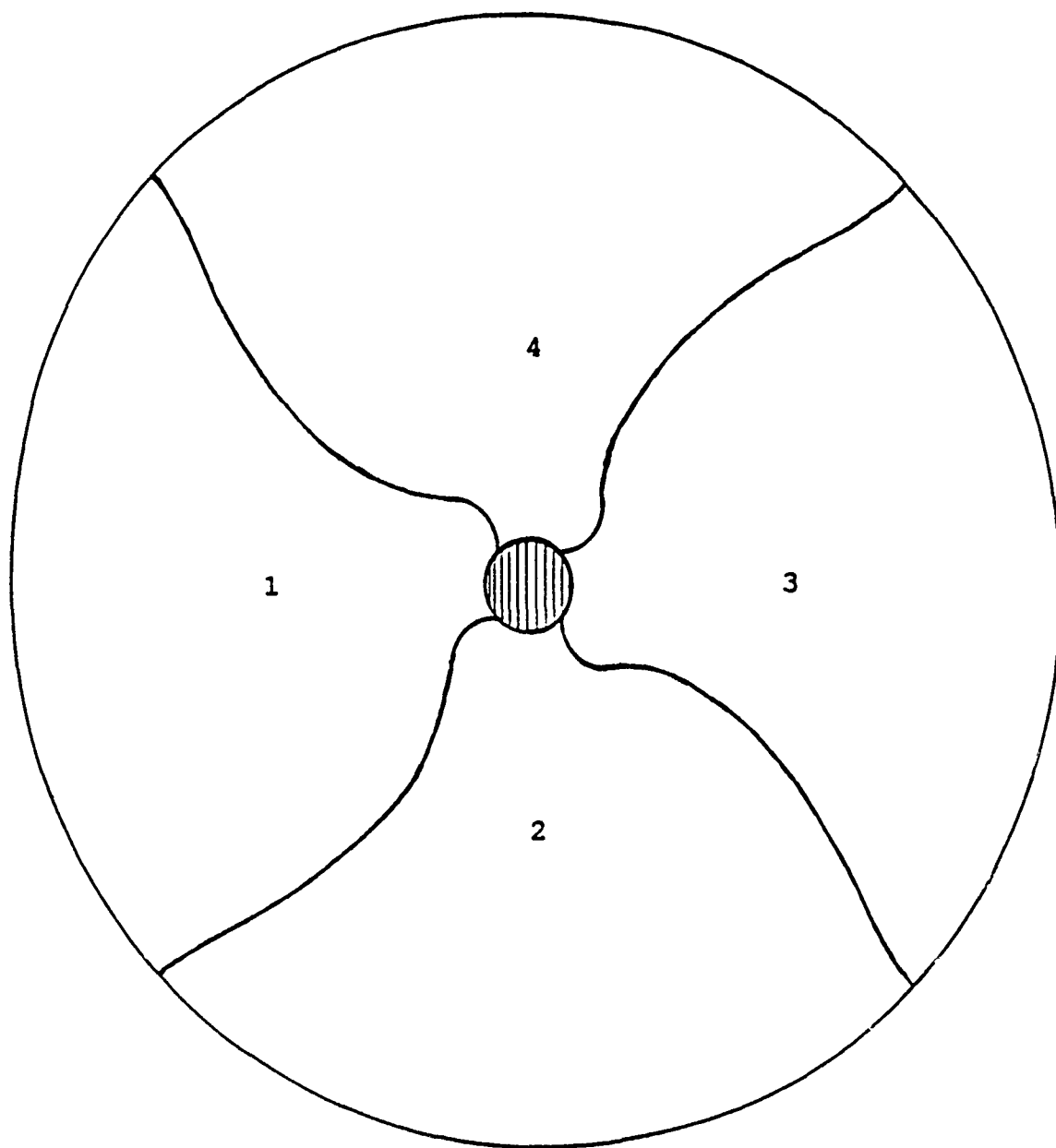


Figure 2. Cross-sectional View at the Middle of the Fins

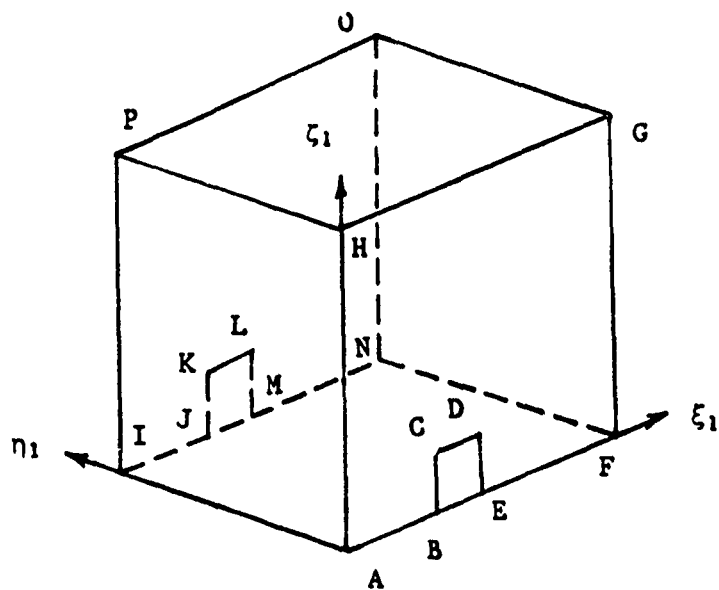
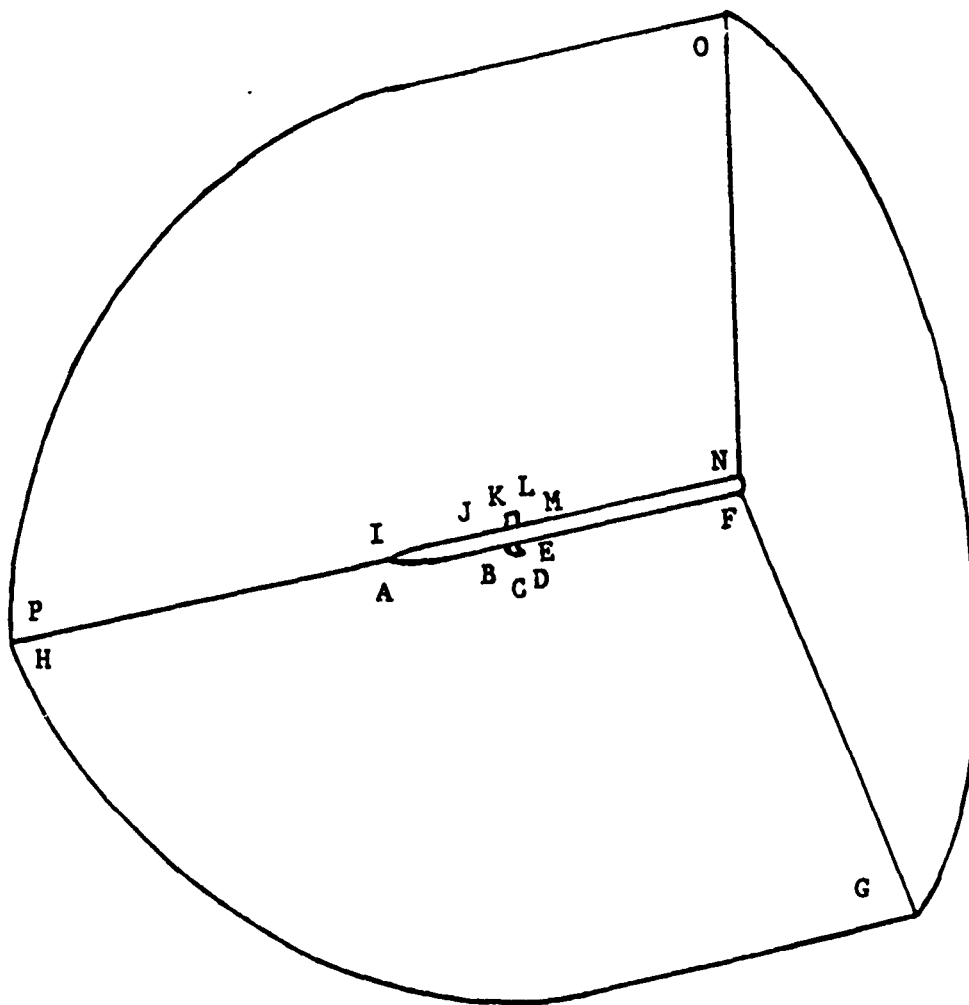


Figure 3. Physical and Computational Space for Block 1

and computational domains for block 1. A careful look at this figure reveals that the transformation used is very straight forward. This is a direct result of the ability of the FRANS code to model more than one no-slip computational surface.

A cut-away view of the three-dimensional grid network developed for the wrap-around fin projectile is shown in Figure 4. This network was generated using the grid generation codes of program EAGLE. An effort was made to pack the grid lines close enough to the missile and fin surfaces in order to resolve the viscous effects in the normal and crosswise directions (see Figure 5). Also, the grid lines in the streamwise direction were clustered at the nose-cylinder junction and at the leading and trailing edges of the fins, in an effort to resolve any shock or expansion phenomena that might appear. Finally, it should be mentioned that the curvature of the grid lines away from the fins was deliberately minimized in order to improve the orthogonal properties of the grid.

GOVERNING EQUATIONS

The unsteady, compressible Navier-Stokes equations without body forces or external heat addition constitute the governing equations for turbulent flow. Unfortunately, these equations are very difficult to solve in their complete and original form, since generally a very large amount of computer time and storage is needed to obtain a solution. Because of this, these equations are generally solved in a time-averaged form known as the Reynolds-averaged Navier-Stokes equations. The Reynolds equations are derived by decomposing the dependent variables in the conservation

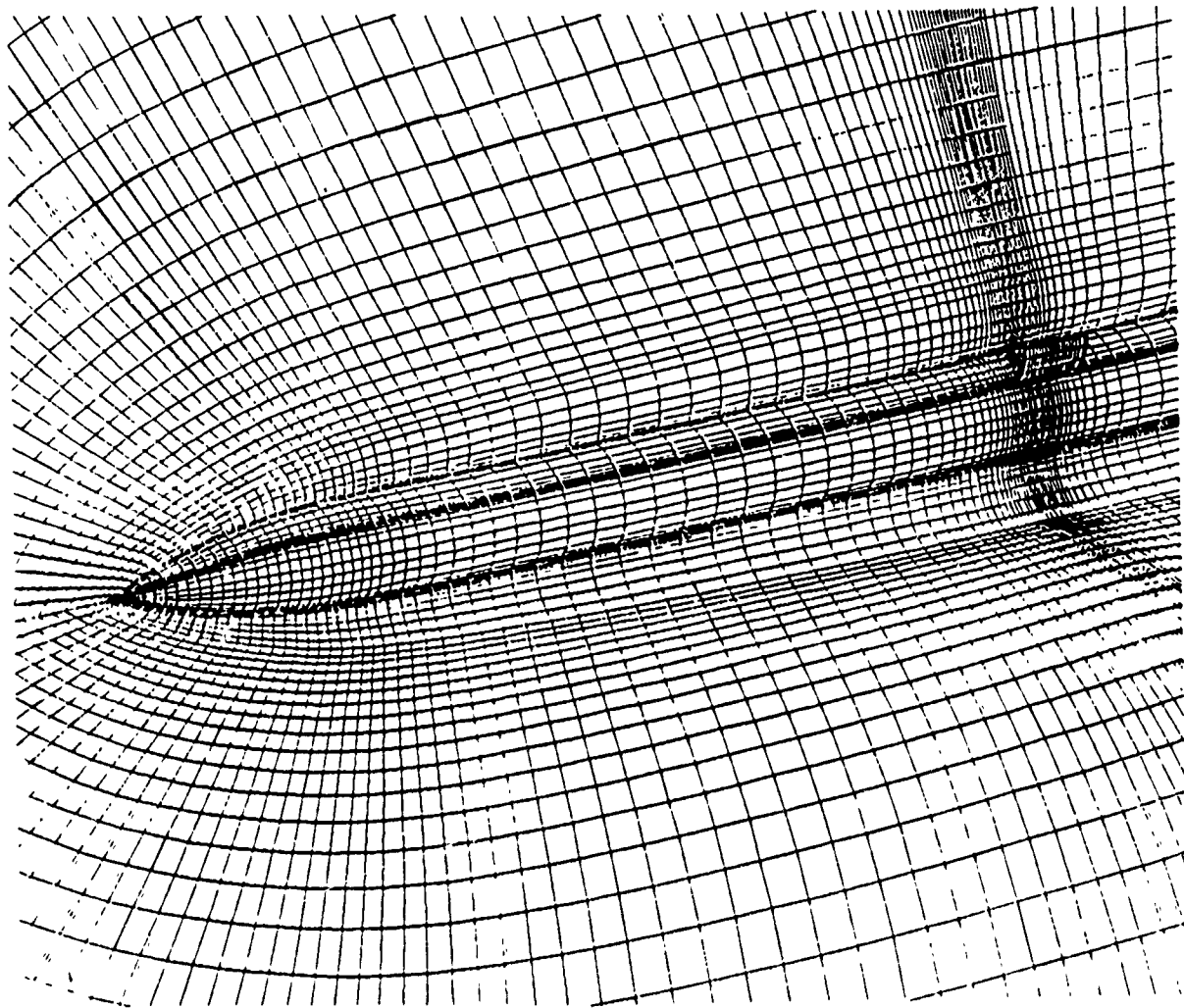


Figure 4. Detail Near the WAF Projectile

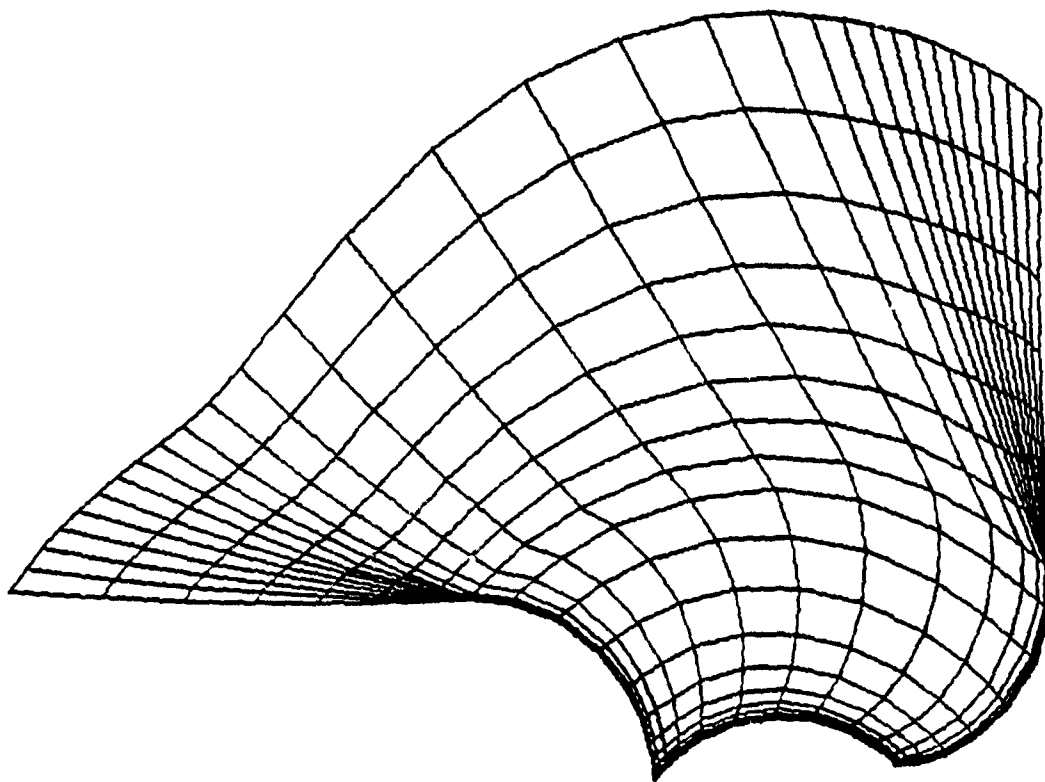


Figure 5. Detail Between Wrap-around Fins

equations into mass-weighted time mean and fluctuating components, and then time averaging the entire equations. This time-averaging yields a series of new terms associated with turbulent motion, which have been interpreted as "apparent" stress gradients and heat-flux quantities. These new quantities constitute new unknowns which must be related to the mean flow variables through turbulence modeling in order to have proper "closure."

Algebraic eddy viscosity models still represent the most popular choice for compressible Navier-Stokes codes, primarily because they are easy to implement and because they require a minimum of computer time and storage. One that is particularly simple is the Baldwin and Lomax model⁴, since it does not require the determination of the boundary-layer thickness or the edge velocity. The turbulent eddy viscosity is computed in two layers using different inner and outer formulations. This model has generally performed well for relatively simple viscous flows, but must be modified in order to handle complicating features.

RESULTS AND DISCUSSION

In order to compute the required value of the eddy viscosity at a particular grid point, the Baldwin and Lomax turbulence model not only requires the perpendicular distance from the point to the solid surface being considered, but it also assumes that certain functional values required in the eddy viscosity formulation are computed on a grid line normal to the surface. Because of this, it was expected that the FRANS code might have difficulty in modeling properly the high grid curvature

regions near the wrap-around fins (recall Figure 5). However, intermediate results of a Mach 0.95 case at an angle of attack of zero degrees suggest that the code is indeed robust enough to handle this complication. Velocity profiles plotted at different planes on the projectile surface as well as on the fin surfaces show that the intermediate solution has the correct qualitative behavior at the solid boundaries. Unfortunately, this particular run eventually resulted in the crash of the program, due to the appearance of a negative value in a square root operation. The error occurred at a point on the stagnation line, very near the nose of the projectile. This suggests that the error may be due to the character of the grid network near the nose.

Under the curvilinear transformation used for this problem, the stagnation line gets mapped onto a singular plane in computational space (recall Figure 3). In theory, the value of the computed solution for any point on a $\zeta = \text{constant}$ line on this singular plane should be the same as that for any other point on that line. However, the value of the Jacobian at each of these points varies tremendously because of the point distribution in the crosswise direction (recall Figure 4). This variation can result in multi-valued "singular" lines, which may eventually lead to the type of error observed. If this is the case, then a simple relaxation of the point clustering near the nose may help to remedy the situation. It should be noted, however, that the code is still in the process of being checked for programming errors. The code has been successfully tested for a flat plate configuration, a transonic airfoil, and a straight-finned missile configuration. Only time and further study will reveal what is the exact nature of the problems being encountered.

CONCLUSION

Intermediate solutions of the Mach 0.95 flow over the wrap-around fin configuration suggest that the FRANS code is robust enough to deal with the complications presented by the high curvature of the grid lines near the fins. Once the current difficulties being experienced with the stagnation region are corrected, the FRANS code promises to be a very effective tool in the numerical solution of the turbulent flow over the wrap-around fin configuration.

ACKNOWLEDGEMENT

This project was conducted under the Air Force Office of Scientific Research (AFOSR) Summer Research Program, at the CFD section of the Air Force Armament Laboratory at Eglin AFB, Florida. Special thanks go out to Nathan Prewitt, Bill Riner, and Magdi Rizk of the Sverdrup Corporation, and to the members of the CFD section for their guidance and assistance. Thanks are also extended to Dr. Bruce Simpson at Eglin and to Dr. Chen-Chi Hsu at the University of Florida for their support and encouragement.

REFERENCES

1. Lijewski, L. E. and Cipolla, J., "Program EAGLE User's Manual," U. S. Air Force Armament Laboratory, Eglin AFB, FL, AFATL-TR-88-117, Vol. I, September 1988.
2. Humphrey, J. A. and Dahlke, C. W., "A Summary of Aerodynamic

Characteristics for Wraparound Fins from Mach 0.3 to 3.0," U. S. Army
Missile Research and Development Command, Redstone Arsenal, AL,
TR-RD-77-5, March 1977.

3. Winchenbach, G. L., Buff, R. S., Whyte, R. H., and Hathaway, W. H.,
"Subsonic and Transonic Aerodynamics of a Wraparound Fin
Configuration," Journal of Guidance, Control, and Dynamics, Vol. 9,
Nov.-Dec. 1986, pp. 627-632.
4. Baldwin, B., and Lomax, H., "Thin Layer Approximation and Algebraic
Model for Separated Turbulent Flows," AIAA Paper 78-257, 1978.

IMPLEMENTATION OF A SYSTEM FOR EVALUATING THE MODULATION TRANSFER FUNCTION OF CCDs USING LASER SPECKLE

**Martin Sensiper - Graduate Research Assistant
(University of Central Florida)**

ABSTRACT:

Many contemporary optical imaging systems contain a focal plane at which the continuous input signal is discretely sampled. In the past, accurately measuring the MTF of such systems has been difficult because the accepted techniques were not designed for the unique problems inherent with sampling. This paper presents a method for establishing the MTF of a sampling array from low frequency to twice Nyquist frequency. The use of random phased laser speckle circumvents the problems inherent with traditional methods of MTF testing where the phase of the test target with respect to the sampling grid affects observed contrast. Another advantage is the simplicity of the equipment and ease with which a test system can be assembled. In this report an aperture is presented which limits the speckle power spectrum to a narrow band, allowing the system response to be evaluated at specific frequencies. The algorithm for extracting the MTF from the data set is described. Sample MTFs are shown using this method and compared with the theoretical MTF and MTFs measured with sine targets.

SECTION I - INTRODUCTION

Modulation transfer function is recognized as the single parameter with which information regarding the total spatial frequency response of an imaging system or subsystem can be described. Widespread use of CCDs has motivated development of an unambiguous method for obtaining their transfer function. To date, the methods for obtaining the MTF curves for optical systems have involved imaging bar targets, sinusoidal targets or knife edges.[References 1,2,3] These deterministic, phase specific test scenes were originally used for deriving the spatial frequency response of continuous imaging systems. When used with imaging systems composed of elemental arrays these methods have serious drawbacks.[Reference 4]

CCDs have discrete elements, pixels, which sample the continuous input. When this input is the image of a test target with a definite phase, the measured response can be a function of the phase of the target image with respect to the phase of the sampling array.[References 4,5,6] A transfer function derived in such a manner is not "shift-invariant". The physical manifestation of this ambiguous transfer function is the imaging of aliased frequency components, those higher than Nyquist frequency.[Reference 5] The Nyquist frequency of a camera is defined as $1/2\Delta x$, where Δx is the pixel center to center spacing. Spatial frequencies higher than the Nyquist frequency, but below the cutoff frequency, will be aliased into lower frequencies. This means

that undersampled high frequencies in the image are reflected symmetrically about the Nyquist frequency, and are displayed as lower frequencies. These aliased frequencies may be considered as edge-images translated into lower frequencies. Generally, it would be of use to know the MTF over the entire spatial frequency spectrum input to the CCD, including those frequencies which may be aliased. While the traditional approaches for finding an MTF generate valid curves at low frequencies, a significant problem remains in that no information about the transfer function past Nyquist frequency is generated.

Previous work has demonstrated the feasibility of using laser speckle as a test target for MTF testing.[References 7,8] The nature of laser speckle insures that the test target phase is random, at any specific point, across the CCD. However, since the statistics of the speckle are well characterized, its average spatial frequency power spectrum, the system input, can be determined. This spectrum can be shaped through use of an aperture. The average spatial frequency power spectrum of the output can be estimated as an ensemble average of the Fourier spectrum of the data from the CCD under test. Since the average power spectrum of the input and output are known, the transfer function can be found.

Lacking in previous demonstrations of the technique, have been appropriate methods for normalization at low frequency. This report concerns implementation of a system to evaluate a specific CCD camera's MTF with laser speckle and the computer software necessary to accomplish this task.

Section III - DEVELOPMENT

The relationship between the intensity distribution of a scattering aperture illuminated by a laser, $|P(x,y)|^2$, and the power spectrum of the speckle at some observation plane, $S(\xi,\eta)$, can be found rigorously through consideration of the second order statistics, namely the autocorrelation, $R(\Delta x, \Delta y)$, of the laser speckle. [Reference 9] To do this, we must first Fourier transform the aperture function to find the mutual intensity of the observed fields and then the autocorrelation of the speckle. Through the Wiener-Khintchine theorem, we can then relate the autocorrelation of the speckle to its power spectrum by another Fourier transform as $\mathcal{F}\{R(\Delta x, \Delta y)\} = S(\xi, \eta)$. Finally, it can be shown that *the power spectrum of the laser speckle has the shape of the normalized autocorrelation of the aperture function plus a delta function at zero frequency.*

The rigorous derivation is neither obvious or intuitive and may be found in Reference 9. A less rigorous explanation is as follows. An aperture is illuminated by frequency coherent, but randomly phased laser radiation. Points in the aperture may be considered as point sources of spherical waves.

On some observation plane, a distance "z" from the aperture, interference patterns caused by the spherical waves combining constructively and destructively may be imaged. The interference is "random" due to the random phase. Points close to one another in the aperture will cause lower spatial frequency interference patterns than those further apart. The "number" of points close or far from each other in the aperture can be found by autocorrelating the aperture.

If the transforms are rigorously performed it can be found that the actual size and location of the features in the power spectrum (in frequency space) are scaled directly by the aperture dimensions and inversely by the wavelength and observation distance as,

$$\xi_n = \frac{L_n}{\lambda z} \quad (1)$$

where ξ_n is a general spatial frequency, L_n is a general aperture dimension, λ is the laser wavelength and z is the distance from the aperture to the observation plane.

We wish to use as an input, laser speckle with a narrow spike at a specific frequency in the power spectrum. The aperture for this is shown in Fig.1. The intensity distribution, $|P(x,y)|^2$, in this aperture can be described as,

$$|P(x, y)|^2 = \text{rect}\left(\frac{x}{l_1}\right) \text{rect}\left(\frac{y}{l_2}\right) * \left[\left(\frac{2}{L}\right) \delta\delta\left(\frac{x}{L/2}\right)\right]. \quad (2)$$

where $*$ is the convolution operator, $\text{rect}(x)=1$ for $|x|\leq 1/2$, zero otherwise, and $\delta\delta=[\delta(x+L)+\delta(x-L)]$, an even-impulse pair.[Reference 10] Following the above formulation relating the aperture function and power spectrum, the autocorrelation of Eq.2, plus the delta function at zero frequency, are equivalent to the speckle power spectrum generated at an observation plane and are given as,

$$S(\xi, \eta) = \langle I \rangle^2 \left[\begin{aligned} &\delta(\xi, \eta) + \frac{1}{2} \frac{(\lambda z)^2}{l_1 l_2} \text{tri}\left[\frac{\xi}{l_1/\lambda z}\right] \text{tri}\left[\frac{\eta}{l_2/\lambda z}\right] \\ &+ \frac{1}{4} \frac{(\lambda z)^2}{l_1 l_2} \text{tri}\left[\frac{\xi-L/\lambda z}{l_1/\lambda z}\right] \text{tri}\left[\frac{\eta}{l_2/\lambda z}\right] \\ &+ \frac{1}{4} \frac{(\lambda z)^2}{l_1 l_2} \text{tri}\left[\frac{\xi+L/\lambda z}{l_1/\lambda z}\right] \text{tri}\left[\frac{\eta}{l_2/\lambda z}\right] \end{aligned} \right] \quad (3)$$

where $\langle I \rangle^2$ indicates an ensemble averaging of speckle intensity, the dimensions l_1, l_2 , and L are given in the aperture (Fig.1 and Eq.2), λ is the laser wavelength, z is the distance from the aperture to the observation plane and $\text{tri}(x)=1-|x|$ for $|x|\leq 1$, zero elsewhere.

This is the average power spectrum input to the focal plane. Note that, in this theoretical construction, half the total power is in the zero frequency delta function and half is in the extended components. If we wish to remove the

delta function, an inverse Fourier transform could be performed, resulting in the autocorrelation of the laser speckle, $\mathcal{F}^{-1}\{S(\xi, \eta)\} = R(\Delta x, \Delta y)$, given as,

$$R(\Delta x, \Delta y) = \langle I \rangle^2 \left[1 + \text{sinc}^2\left(l_1 \frac{\Delta x}{\lambda z}\right) \text{sinc}^2\left(l_2 \frac{\Delta y}{\lambda z}\right) \cos^2\left(2\pi L/2 \frac{\Delta x}{\lambda z}\right) \right] \quad (4)$$

The constant first term, a DC offset, can be subtracted off in this domain. When we retransform this function, $\mathcal{F}\{R_{\text{normalized}}(\Delta x, \Delta y)\}$, we have Eq.3 without the delta function. We can determine the attenuation of the modulated triangles since the zero frequency triangle is not attenuated.

The function in Fig.1 and Eqs.2 and 3 is separable in x and y . Information about the power spectrum along the horizontal frequency axis of the two-dimensional transform, Eq.3, is available as the power spectrum, modulus squared of a one-dimensional Fourier transform, of the laser speckle intensity along horizontal spatial axis. We can therefore reduce Eq.3 to,

$$S(\xi) = \langle I \rangle^2 \left[\begin{aligned} &\delta(\xi) + \frac{1}{2} \frac{\lambda z}{l_1} \text{tri}\left[\frac{\xi}{l_1/\lambda z}\right] \\ &+ \frac{1}{4} \frac{(\lambda z)}{l_1} \text{tri}\left[\frac{\xi - L/\lambda z}{l_1/\lambda z}\right] \\ &+ \frac{1}{4} \frac{(\lambda z)}{l_1} \text{tri}\left[\frac{\xi + L/\lambda z}{l_1/\lambda z}\right] \end{aligned} \right] \quad (5)$$

Fig.2 shows this one dimensional "slice" of the two-dimensional power spectrum along the ξ axis. The features in the power spectrum scale with z as indicated by Eq.1. As z increases, the modulated triangles will "migrate"

toward zero frequency. In the speckle domain, this is equivalent to the size of the speckle increasing as the observation plane moves away from the aperture. We can therefore control the exact modulation frequency of the speckle incident on the array. Lines or columns of pixels of a focal plane array placed to observe the speckle, generate samples of the speckle intensity. Each line or column of data can be interpreted as a single observation of a ergodic random process. We obtain an estimate of the true power spectrum through ensemble averaging the squared modulus of the Fourier transform of the intensity. The exact procedure used to extract the system response is described in Sec.5.

SECTION IV TEST INSTRUMENT

The equipment necessary to implement this procedure is shown schematically in Fig.3. A complete listing of the actual components used can be found in Appendix 1. Each component is necessary to fulfill certain theoretical assumptions or requirements, and to create the desired speckle test pattern.

These assumptions include:

1. Monochromatic light.
2. Uniform distribution of the phase of the electric field in the aperture.
3. Uniform illumination of the aperture.
4. Propagation of one linearly polarized component.

Assumption one implies that the laser emission is limited to a narrow linewidth. Many lasers meet this requirement, however, the inherent inefficiency of the integrating sphere restricts the choice of lasers. Enough optical power over the narrow laser linewidth must be input to overcome the loss of power seen at the output port. The characteristics of laser speckle generated by an integrating sphere were investigated in Reference 7 and shown to meet requirements two and three. Monitoring the camera output, while adjusting the laser power level, assures that the contrast in the speckle provides as great a dynamic range as possible. Notice that as the camera is moved away from the aperture, the flux will drop toward a point where there is no contrast. Inclusion of a polarizer in the optical path assures that assumption four is met.

SECTION V - EXPERIMENTAL PROCEDURE

The data processing relies on performing numerous fast Fourier transforms (FFTs) on speckle images and then averaging the results in frequency space. Before the FFTs are done, histograms of each frame are examined. This is done to insure that the speckle intensity probability distribution function is appropriately shaped.[References 7,9] The histogram also indicates if excess flux has saturated the CCD. If a high percentage of pixels have values of 255, this is assumed to be the case and the sample is discarded.

Note that power in the Nyquist frequency is referred to the highest frequency component of the FFT. A ratio can thus be formed relating spatial frequencies in the speckle power spectrum to each of n^{th} components in the FFT. Therefore,

$$\frac{\text{Nyquist frequency}}{\text{highest numbered frequency component}} = \frac{\text{spatial frequency}}{n^{\text{th}} \text{ component}} \quad (6)$$

Using Eq.6, values can then be found for spatial frequencies, ξ_n , corresponding to each FFT component, n . Recall that an FFT of $2n$ samples has as an output n frequency components. For frequencies from Nyquist to twice Nyquist, n is incremented from n to $2n$. Eq.1 is then solved for each distance z_n , so the position for the observation plane which places the modulated triangles, Fig.2, at specific frequencies can be established.

This procedure, to first calculate the desired imaging distances and place the observation plane at these distances corresponding to exact FFT frequencies, was used for the first of the two data sets, Fig.4. When each frame of data is reduced to a single average power spectrum, the value of the maximum of the shifted triangle is extracted. This is a simpler algorithm if the exact component at which it occurs is known. Secondly, this might minimize leakage due to the discrete transform process. For the second data set, Fig.5, distances were chosen randomly and the peak of the triangle was found by the computer.

One-dimensional FFTs of each speckle image were performed on a line by line basis. The squared moduli of the results are averaged, creating an ensemble average of the power spectral density, (PSD), along the ξ axis. The DC bias in this domain is a large delta function at zero frequency. (See Eq.3) Recall that half the power is in this delta function for the theoretical input. In the actual data, additional power is in this delta function from several sources. Ambient light from sources other than the aperture, or laser light reflecting from nearby surfaces adds to the DC background. We wish to remove this DC, but to retain the triangle at zero frequency in order that the modulated triangles, attenuated by the CCD MTF, can be properly scaled. Note, in Fig.2 and Eqs.3 and 5, that in the input spectrum, the modulated triangles are one-half the height of the zero frequency triangle.

To eliminate the delta function, we inverse transform the average PSD and obtain an average autocorrelation as shown in Eq.4. We subtract off the minimum value, a DC offset, then retransformed the result. The modulated triangle is attenuated by the CCD MTF and the modulated triangle's peak represents the output power spectrum at a particular frequency. Recall that the MTF is the magnitude of the optical system transfer function. As developed in linear systems theory, the relationship between output power spectral density $S_{out}(\xi)$, input power spectral density $S_{in}(\xi)$ and a system transfer function $H(\xi)$ are related by,

$$S_{out}(\xi) = |H(\xi)|^2 S_{in}(\xi). \quad (7)$$

Using the ratio of 2:1, zero frequency triangle to modulated triangle in the input spectrum, we can find the amount that the output power spectrum was attenuated and take the square root of this to generate one point on the MTF curve. Systematically doing this with speckle images taken at various distances and relating the spatial frequency of the triangle's peak through Eq.1, we can generate a MTF curve from zero frequency to near Nyquist frequency. For frequencies beyond Nyquist, the modulated frequency triangle is aliased symmetrically about Nyquist into lower frequencies. When plotting the MTF curve these data points are simply folded out into the actual frequency of input. The high frequency limit is the overlap of the modulated triangle - aliased into lower and lower frequencies - with the unmodulated triangle, as we approach twice Nyquist frequency.

The algorithm for finding the points on the PSD is reiterated here for clarity.

1. Perform 1-D Fourier transform of lines/columns of sampled speckle data.
2. Find the power spectrum = squared modulus of the Fourier transform.
3. Average the power spectra of all the lines/columns in an image.
4. Inverse transform the average power spectrum = average autocorrelation.
5. Subtract off minimum value = normalized autocorrelation.

6. Retransform normalized autocorrelation = normalized power spectrum.
7. Divide normalized power spectrum by value at zero frequency normalizing zero frequency peak to 1.
8. Multiply value of the peak of modulated triangle by 2 to scale vs. zero frequency triangle and take the square root of this value = one point on MTF curve.
9. Do this for numerous distances which can be related to spatial frequencies by Eq.1.

It should be noted that computer code written to find the local maximum where the peak of the modulated triangle occurs, depends on only approximately knowing where the peak is. This approximate distance for each image is determined by measuring the aperture to image plane distance as accurately as possible while taking the data. The computer program then searches the local region near the FFT components defined by the approximate distance, number of pixels per line/column, their width, the laser wavelength and the aperture dimensions. The system is therefore effectively "self-calibrated".

SECTION VI - EXPERIMENTAL RESULTS

Due to time limitations, only two complete data sets were processed and while they demonstrate that the system works qualitatively, they contain indications

of that more work needs to be done. The resulting series of PSDs are most dramatically viewed as a "movie" or slide show. These are not included here for obvious reason of space. As z decreases, the modulated triangle travels from low to high spatial frequency, "bounces" at the Nyquist frequency, and proceeds back toward zero frequency. These triangles, appropriately scaled as discussed above, plot the MTF curve, examples of which are seen in Figs. 4 and 5. A curve was fit to the data which was appropriate.

Note in Fig.4, data set 1, that a values around Nyquist are very high and were therefore left out of the curve fit. Both data sets end with a point left out of the curve. This point occurred as the modulated triangle started to overlap the unmodulated triangle and is invalid. In both data sets, the MTF at low frequencies drops off. Both the response at Nyquist and the low frequency attenuation need to be explained by collecting more data.

Where this system has been implemented previously, the DC gain and offset of the frame grabber were accessible. This was not the case here and this could be part of the problem. With no input to the camera, the frame grabber output was all zeros. Ordinarily, the offset should be set so that some small positive values of noise are seen in all pixels. This is then removed by direct subtraction of this "dark" frame or by subtraction of the noise power spectrum.

The theoretical MTF for a rectangular pixel is given functionally as $\sin(\pi\xi\Delta x)/(\pi\xi\Delta x)$, where Δx is the pixel pitch in the appropriate direction. It is not expected nor even desired that the response curve approach this theoretical curve. This function has a value of ~63% at Nyquist frequency and is shown in Fig.6. Figure 6 also contains a curve fit to data obtained by imaging a target with sinusoids of various frequencies. Video cameras, such as the Pulnix TM745, contain electronic filters which intentionally degrade the response around the Nyquist frequency. While this reduces aliasing, the tradeoff is a limiting of response below the Nyquist frequency. Again, I believe good speckle data would show this.

The speckle data does not match the theoretical curve, which is as expected, nor does seem to correlate well with the data obtained with the sinusoidal test target. However, in the limited time available, the software necessary to perform the desired operations was completed, tested, and documentation provided, sufficient for future users to expand upon. Complete copies of the programs, hardcopy of the code, along with step by step explanation of the code was provided.

SECTION VII - Conclusions

The use of laser speckle for measuring the MTF of sampling arrays to twice Nyquist frequency has been demonstrated. The results are promising but

leave work to be done. Much better results are possible and have been obtained in other implementations of this system. The advantages in obtaining a MTF curve in this fashion should again be noted.

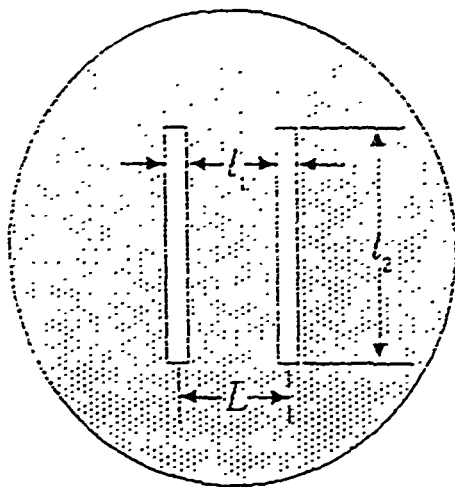
The components are readily available or easy to fabricate, plus the test bed can be set up with minimal calibration. Since the optical path contains no "imaging" optics, their cost or contribution to the overall MTF does not need to be considered. In addition, the use of this method for testing systems at wavelengths inaccessible by prior procedures should be considered. The method is, in effect, self calibrating. The spatial frequencies of interest, as well as the MTF at those frequencies, are found by the computer program as described.

Further work needs to be done, particularly with regard to the digitizer/camera interface. The possibility of data reduction by binarizing the data set is another interesting variation to be tried. Finally, a complete computer controlled system with a motorized stage to move the camera and automated data acquisition might be implemented. The method is computer intensive. However, the widespread availability of efficient image processing software and hardware will eliminate this as a drawback. If efficient techniques are to be developed to eliminate or use the spatial frequency components above Nyquist, an exact function for the system response needs to be determined.

Appendix 1

This is a list of the components used in the Special Projects Lab at Wright Lab, Eglin A.F.B., Summer 1991, to generate and process laser speckle for testing the Pulnix 745E.

Laser	Spectra Physics 165 Laser - Argon laser with etalon adjusted for $\lambda = 514 \text{ nm}$.
Integrating Sphere	1" Labsphere IS-010-SF with Spectraflex(barium sulphate) coating.
Polarizer	Ambico R-8945.
Aperture	Custom aperture currently at UCF/CREOL. (See Fig.1. $L = 0.06"$, $l_1 = 0.005"$, $l_2 = 0.08"$)
Camera	Pulnix 745E, with AGC = off and gamma = 1.
Computer	The IPL/RSPL computer facility was used, specifically the Gould/DiAnza IP8500 Digitizer. The computer programs necessary to process the data were written in VAX Fortran. The programs and the data are on tape and are completely documented in a separate document, a copy of which was left with the lab personnel.



Aperture Geometry
for
Narrowband
Power Spectrum

Fig.1

1-D Power Spectrum
Along ξ axis
with Narrowband Aperture

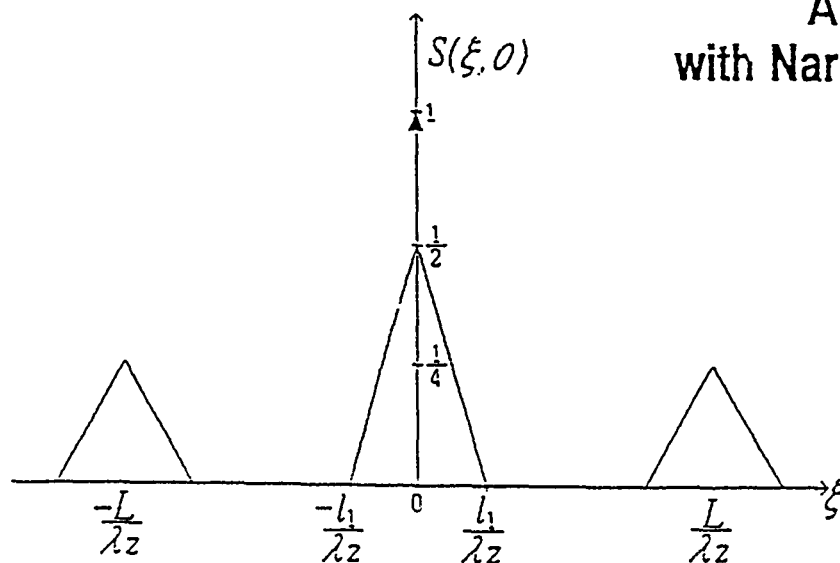


Fig.2

Block Diagram of Speckle Generator

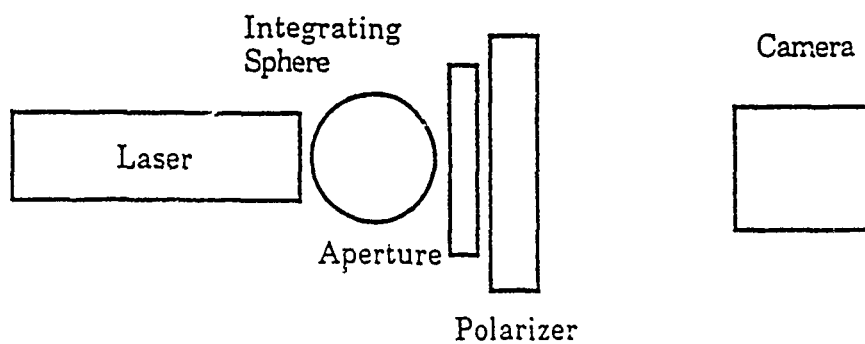


Fig.3

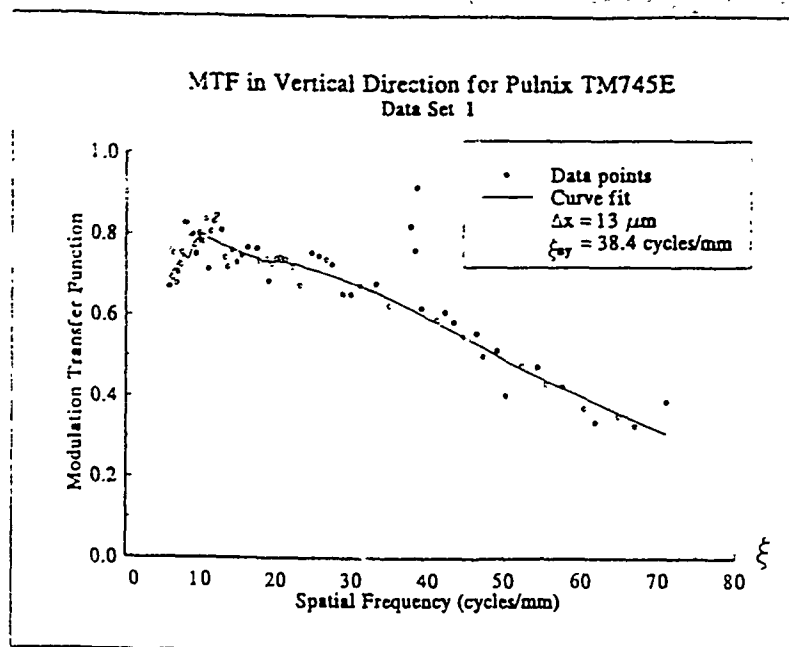


Fig.4

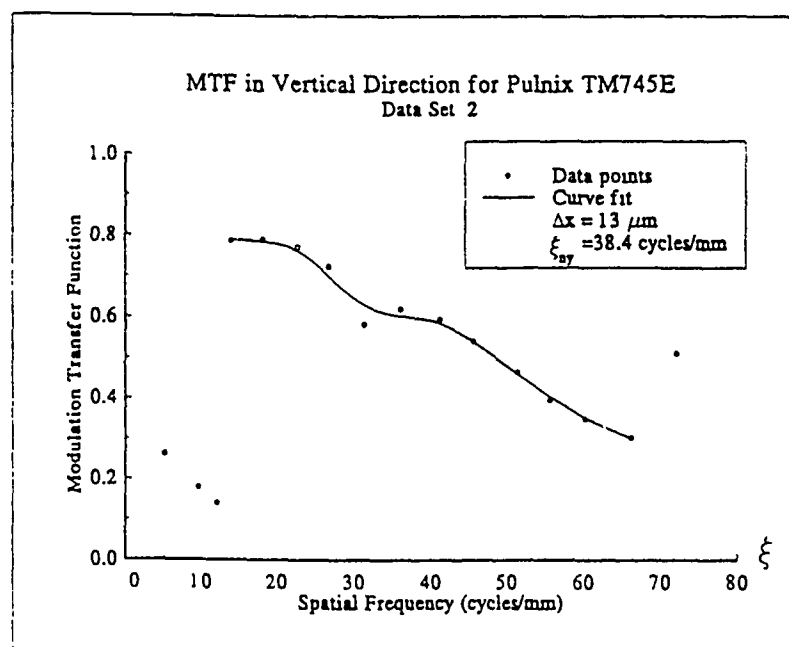


Fig.5

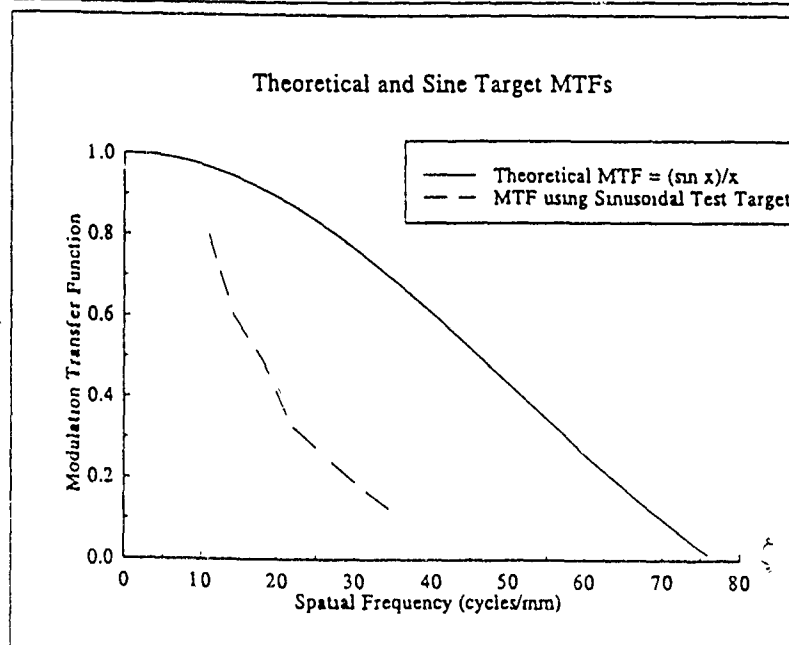


Fig.6

Bibliography

1. S.K.Park, R.Schowengerdt and M.Kaczynski, "Modulation-transfer-function analysis for sampled image systems," *Appl. Opt.* 23(15), 2572-2582(1984).
2. S.B.Campana, "Techniques for Evaluating Charge Coupled Imagers," *Opt. Eng.* 16(3), 2267-2274(1977).
3. S.E.Reichenbach, S.K.Park and R.Narayanswamy, "Characterizing digital image acquisition devices," *Opt. Eng.* 30(2), 170-177(1991).
4. W.Wittenstein, et al., "The definition of the OTF and the measurement of aliasing for sampled imaging systems", *Opt. Acta* 29(1), 41-50(1982).
5. R.Legault, in *Perception of Displayed Information*, L.M.Biberman, ed., pp.279-312, Plenum, New York (1973)
6. J.Feltz and M.Karim, "Modulation transfer function of charge-coupled devices," *Appl.Opt.* 29(5), 717-722(1990).
7. G.Boreman, Y.Sun and A.James, "Generation of Laser Speckle with an Integrating Sphere," *Opt.Eng.* 29(4), 339-342(1990).
8. G.Boreman and E.Dereniak, "Method for measuring modulation transfer function of charged-coupled devices using laser speckle," *Opt.Eng.* 24(1), 148-150(1986).
9. J.Goodman, in *Laser Speckle and Related Phenomena*, J.C.Dainty, ed., pp. 35-40, Springer-Verlag, Berlin(1975).
10. J.Gaskill, *Linear Systems, Fourier Transforms, and Optics*, p.57, Wiley, New York(1978)

**1991 USAF-UES SUMMER FACULTY RESEARCH PROGRAM/
GRADUATE STUDENT SUMMER SUPPORT PROGRAM**

**Sponsored by the
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH**

**Conducted by the
Research Development Laboratories**

FINAL REPORT

A NOTE ON PRONY'S METHOD

Prepared by:	David Choate and Wendy Barkman
Academic Rank:	Associate Professor
Department and	Science Division
University	Transylvania University
Research Location:	Air Force Avionics Laboratory, Electronic Support Measure Group
USAF Researcher:	Dr. Jim Tsui
Date:	August 12, 1991
Employee Number:	183

A NOTE ON PRONY'S METHOD

David B. Choate, Associate Professor, and Wendy Barkman

ABSTRACT- Prony's method can be used to compute the frequencies of two signals using four receivers. It is shown in this note that this can be done with only three receivers.

I. INTRODUCTION

Briefly stated Prony's algorithm [1] for the two signal case can be presented as follows:

Assume that the unknown frequencies, a and b , are roots of the quadratic equation

$$y^2 - \alpha_1 y - \alpha_2 = 0.$$

Given the four observation equations from receivers

- (i) $x_0 = c_1 + c_2$
- (ii) $x_1 = c_1 a + c_2 b$
- (iii) $x_2 = c_1 a^2 + c_2 b^2$
- (iv) $x_3 = c_1 a^3 + c_2 b^3$

, we first multiply equation (i) by α_2 , equation (ii) by α_1 and equation (iii) by -1 to obtain

$$\alpha_2 x_0 = c_1 \alpha_2 + c_2 \alpha_2$$

$$\alpha_1 x_1 = c_1 \alpha_1 a + c_2 \alpha_1 b$$

$$-x_2 = -c_1 a^2 - c_2 b^2 .$$

Adding these three equations gives

$$\begin{aligned} \alpha_2 x_0 + \alpha_1 x_1 - x_2 &= c_1(\alpha_2 + \alpha_1 a - a^2) + c_2(\alpha_2 + \alpha_1 b - b^2) \\ &= 0 + 0 = 0 , \text{ or} \end{aligned}$$

$$\alpha_1 x_1 + \alpha_2 x_0 = x_2 \quad (1)$$

Now multiply equation (ii) by α_2 , equation (iii) by α_1 and equation (iv) by -1 to obtain

$$\alpha_1 x_2 + \alpha_2 x_1 = x_3 \quad (2) .$$

Equations (1) and (2) yield solutions for the α 's provided

$$(x_1)^2 \neq x_0 x_2 . \quad (3)$$

Substituting these values into the original quadratic gives

$$a, b = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

But observe that equation (3) demonstrates that the existence of distinct solutions depends only on the results of the first three receivers: x_0 , x_1 and x_2 . So it is reasonable to ask whether their location depends only on just these x 's as well.

Furthermore, a and b are typically assumed to be on the

unit circle. And yet nowhere in Prony's derivation is this property exploited. In this paper we will show that if attention is paid to this detail only three receivers are required to obtain both solutions.

II. THREE-FOUR

Consider the system

$$\begin{aligned} (i) \quad x_0 &= c_1 + c_2 \\ (ii) \quad x_1 &= c_1 e^{-j\theta_1} + c_2 e^{-j\theta_2} \\ (iii) \quad x_2 &= c_1 e^{-j2\theta_1} + c_2 e^{-j2\theta_2} \end{aligned} \quad (I)$$

, where the c 's are nonzero complex numbers and the θ 's are real. Both are unknown. And the x 's are given complex numbers.

Set $a = e^{-j\theta_1}$ and $b = e^{-j\theta_2}$ and assume that $a \neq b$. System (I) is equivalent to

$$\begin{pmatrix} 1 & 1 \\ a & b \\ a^2 & b^2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

which in turn is equivalent to

$$\begin{pmatrix} 1 & 1 & 1 \\ a & b & 0 \\ a^2 & b^2 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \quad (II)$$

(Note that the 1 we have added to the 1x3 position can be replaced by any nonzero constant without changing the sequel.)

Since $a \neq b$, the inverse of the coefficient matrix of (II) is

$$\begin{pmatrix} 0 & -b/(a^2-ab) & 1/(a^2-ab) \\ 0 & (a^2/b)/(a^2-ab) & (-a/b)/(a^2-ab) \\ 1 & [b-(a^2/b)]/(a^2-ab) & -[1-(a/b)]/(a^2-ab) \end{pmatrix}$$

After solving the system we obtain

$$c_1 = [-b/(a^2-ab)]x_1 + [1/(a^2-ab)]x_2$$

$$c_2 = [(a^2/b)/(a^2-ab)]x_1 + [(-a/b)/(a^2-ab)]x_2$$

By (I) i we get

$$x_0 = c_1 + c_2 = \left[\frac{a+b}{ab}\right]x_1 + \left[\frac{-1}{ab}\right]x_2.$$

$$\text{So } abx_0 = (a + b)x_1 - x_2. \quad (3)$$

Since a is on the unit circle, $1/a = a^*$, the complex conjugate of a . Taking conjugates of (3) gives

$$ab(x_2)^* = (a + b)(x_1)^* - (x_0)^* . \quad (4)$$

Solving for a in both (3) and (4) yields

$$\frac{bx_1 - x_2}{bx_0 - x_1} = \frac{b(x_1)^* - (x_2)^*}{b(x_2)^* - (x_1)^*}$$

provided $b \neq x_1/x_0$.

Expanding and combining gives

$$[x_1(x_2)^* - (x_1)^*x_0]b^2 + [x_0(x_0)^* - x_2(x_2)^*]b + [x_2(x_1)^* - (x_0)^*x_1] = 0 .$$

$$\text{Set } A = x_1(x_2)^* - (x_1)^*x_0$$

$$B = x_0(x_0)^* - x_2(x_2)^* = |x_0|^2 - |x_2|^2$$

$$\text{and } C = A^* = x_2(x_1)^* - (x_0)^*x_1 .$$

Therefore

$$e^{-j\theta_2} = b = \frac{B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

provided $A \neq 0$.

But (3) and (4) are symmetric in a and b. So by a similar computation

$$e^{j\theta_1} = a = \frac{-B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

provided $A \neq 0$.

But since $a \neq b$ we conclude after checking in system (I) that

$$\theta_1 = \left\{ \ln \left(\frac{-B - \sqrt{B^2 - 4|A|^2}}{2A} \right) \right\} j$$

and

$$\theta_2 = \left\{ \ln \left(\frac{-B + \sqrt{B^2 - 4|A|^2}}{2A} \right) \right\} j$$

But certainly there is something disturbing about this derivation. In completing it we have assumed that $|a| = |b| = 1$. And yet the expression

$$\frac{-B \pm \sqrt{B^2 - 4|A|^2}}{2A}$$

will have a norm of 1 only if $B^2 - 4|A|^2 \leq 0$.

So at first it would seem that three receivers will suffice only if this inequality is satisfied. But we can avoid this restriction by a slight of hand: we will simply abandon our requirement that a and b lie on the unit circle as soon as it cannot be satisfied.

When $B^2 - 4|A|^2 > 0$, set

$$\theta_1 = \beta_1 + \gamma_1 j, \text{ its complex form.}$$

Then

$$a = e^{-j\theta_1} = e^{\gamma_1} e^{-j\theta_1}$$

Thus one frequency is the real part of θ_1 and, of course, the other is the real part of θ_2 . These values appear to agree with Prony's- at least to fifteen decimal places.

But there are still two cases for which we have not obtained a solution; namely, $b = x_1/x_0$ and $A = 0$.

III. SHUT THE DOOR

To deal with each of these cases we will examine the polar form of our x 's and c 's:

$$x_i = |x_i| e^{j\alpha_i}, \quad i = 0, 1, 2$$

$$c_i = |c_i| e^{j\phi_i}, \quad i = 1, 2$$

With these equations we will find that the case $b = x_1/x_0$ is not really a possibility under the assumption that $a \neq b$.

Case 1: If the given x 's are nonzero, then $a = b$ only if $b = x_1/x_0$.

$$\begin{aligned} \text{Proof. } (\rightarrow) \text{ If } a = b, \text{ then } x_1 &= c_1 a + c_2 b \\ &= (c_1 + c_2) b \\ &= x_0 b \end{aligned}$$

(\leftarrow) Suppose $b = x_1/x_0$. If $a \neq b$, then by equation (3)

$$0 = a(bx_0 - x_1) = bx_1 - x_2$$

which implies that $b = x_2/x_1$. So $(x_1)^2 = x_0x_2$. Therefore

$$\alpha_1 = (\alpha_0 + \alpha_2)/2 \quad (5)$$

Furthermore $b = (x_1)^*/(x_2)^*$. Thus

$$\begin{aligned} 0 &= A = x_1(x_2)^* - (x_1)^*x_0 \\ &= (c_1a + c_2b) \{ [(c_1)^*/a^2] + [(c_2)^*/b^2] \} \\ &\quad - \{ [(c_1)^*/a] + [(c_2)^*/b] \} (c_1 + c_2), \text{ or} \end{aligned}$$

$$a^2 = \left[\frac{c_2(c_1)^*}{c_1(c_2)^*} \right] b^2$$

Thus by (I)iii

$$\begin{aligned} x_2 &= \left[\frac{c_2}{(c_2)^*} \right] \left[\frac{(c_1)^* + (c_2)^*}{(c_2)^*} \right] b^2 \\ ; \text{ so } \frac{x_2}{(x_0)^*} &= \left[\frac{c_2}{(c_2)^*} \right] b^2 \end{aligned}$$

Taking logs and using the polar forms of x and c gives

$$\alpha_1 = \phi_2 - \theta_2.$$

Similarly,

$$\alpha_1 = \phi_1 - \theta_1.$$

So, of course,

$$\phi_2 - \phi_1 = \theta_2 - \theta_1.$$

By (I)ii we obtain:

$$\begin{aligned}
|x_1|e^{a_1} &= |c_1|e^{j\phi_1}e^{-j\theta_1} + |c_2|e^{j\phi_2}e^{-j\theta_2} \\
&= |c_1|e^{ja_1} + |c_2|e^{ja_1}, \text{ or} \\
|x_1| &= |c_1| + |c_2|. \quad (6)
\end{aligned}$$

But since $b = x_1/x_0$, $|x_0| = |x_1|$

(7) So

$$|c_1 + c_2| = |c_1| + |c_2|.$$

By the triangle inequality $\phi_1 = \phi_2$, or $\theta_1 = \theta_2$.

So $a = b$, a contradiction.

Case 2: If $A = 0$ and the x 's are nonzero, then a solution to the system (I) is given by

$$\begin{aligned}
a &= \frac{x_2}{x_1} - \sqrt{\frac{x_2^2}{x_1^2} - \frac{x_2}{x_0}} \\
&\text{and} \\
b &= \frac{x_2}{x_1} + \sqrt{\frac{x_2^2}{x_1^2} - \frac{x_2}{x_0}}.
\end{aligned}$$

Proof. From the information gathered in the proof of Lemma 1 we have

$$\begin{aligned}
x_2 &= c_1 a^2 + c_2 b^2 \\
&= |c_1|e^{ja_1}a + |c_2|e^{ja_1}b, \text{ or} \\
\frac{|x_1|x_2}{x_1} &= |c_1|a + |c_2|b \quad (8)
\end{aligned}$$

After writing (6) and (8) in matrix form we get

$$\begin{pmatrix} 1 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} |c_1| \\ |c_2| \end{pmatrix} = |x_1| \begin{pmatrix} 1 \\ \frac{x_2}{x_1} \end{pmatrix}$$

from which we obtain

$$|c_1| = \frac{|x_1|}{(b-a)} \left[b - \frac{x_2}{x_1} \right] \quad (9)$$

$$|c_2| = \frac{|x_2|}{(b-a)} \left[-a + \frac{x_2}{x_1} \right] \quad (10)$$

Now suppose $|c_1| = |c_2|$.

Using (7), (9) and (10) we get

$$a + b = 2x_2/x_1 \quad (11)$$

But by (3) and (11)

$$abx_0 = 2x_2 - x_2 = x_2 ,$$

$$\text{or } ab = x_2/x_0. \quad (12)$$

Solving (11) and (12) simultaneously gives our formulas.

Our above argument certainly provides solutions, but it does not demonstrate that they are unique. However, our formulas yield frequencies which agree with Prony's method.

So it appears we can determine our frequencies in this situation unless it so happens that an x is zero. But as we

might expect deriving solutions for degenerate spectra amounts to little more than manual labor.

IV. PICK UP STICKS

If all our x 's are 0, then there is only one signal and it is not unique.

If $x_1 = x_2 = 0$, then so is x_3 . So, again, there is only one signal and it is not unique. This also follows if $x_1 = x_2 = 0$.

But if $x_0 = x_2 = 0$ and $x_1 \neq 0$, then $\theta_1 = 3\pi/2$ and $\theta_2 = \pi/2$.

If $x_1 = 0$, $x_0 \neq 0$ and $x_2 \neq 0$, then

$$\theta_1 = (\alpha_1 - \alpha_2)/2$$

and

$$\theta_2 = \theta_1 + \pi.$$

If $x_2 = 0$ and $x_0 \neq 0 \neq x_1$, then $A \neq 0$ and we can use our derived formulas.

If $x_0 = 0$ and $x_1 \neq 0 \neq x_2$, then $A \neq 0$ in this case as well.

So the solutions to system (I) have been determined.

V. LAY THEM STRAIGHT

Finally it would be convenient for the sake of programming to be able to determine at the outset whether the

signals are identical. To this end we prove:

Theorem. For nonzero x the equation $a = b$ holds only if $|x_0| = |x_1| = |x_2|$.

Proof. (\leftarrow) Since $a = b$, $x_1 = (c_1 + c_2)a = x_0a$, or $x_1/x_0 = a$.

Thus

$$|x_0| = |x_1|. \text{ Similarly } |x_1| = |x_2|.$$

(\rightarrow) If $|x_0| = |x_1|$, then $|c_1 + c_2| = |c_1| + |c_2|$. By the triangle inequality $\phi_1 = \phi_2$ which implies $\theta_1 = \theta_2$, or $a = b$.

REFERENCES

1. Hildebrand, F. B., Introduction to Numerical Analysis, McGraw-Hill, New York, 2nd ed., 1956.

APPLYING CASCADE CORRELATION TO THE EXTRAPOLATION OF CHAOTIC TIME SERIES

David Ensley
Department of Electrical Engineering
Auburn University
September 1991

Abstract

Attempting to find near-optimal architectures, ontogenic neural networks develop their own architectures as they train. As part of a project entitled "Ontogenic Neural Networks for the Prediction of Chaotic Time Series," this paper presents findings of a ten-week research period on using the Cascade Correlation ontogenic neural network to extrapolate (predict) a chaotic time series generated from the Mackey-Glass equation. Truer, more informative measures of extrapolation accuracy than currently popular measures are presented. The effects of some network parameters on extrapolation accuracy were investigated. Sinusoidal activation functions turned out to be best for our data set. The best range for sigmoidal activation functions was $[-1, +1]$. One experiment demonstrates that extrapolation accuracy can be maximized by selecting the proper number of training exemplars. Though surprisingly good extrapolations have been obtained, there remain pitfalls. These pitfalls are discussed along with possible methods for avoiding them.

Introduction

The project entitled "Ontogenic Neural Networks for the Prediction of Chaotic Time Series," of which this work is a part, has two primary motivations. One is to explore and expand the relatively new genre of neural networks called ontogenic neural networks. The name *ontogenic* comes from the biological term *ontogenesis*, which *The American College Dictionary* [Barnhart, 1970] defines as "the development of an individual organism." Thus, ontogenic neural networks develop their architectures as they learn instead of requiring the user to specify an architecture before training.

Traditionally neural network developers have had to find the optimum architecture for their networks by trial and error. In the last three years or so, researchers have made progress in developing networks that attempt to determine the optimum architecture automatically. Alpaydm recently published

a taxonomy of such networks he calls incremental learning [Alpaydin, 1991]. His two classes of incremental learning are networks that start with many nodes, pruning themselves as they train, and networks that start with a minimum number of nodes, adding more as they train. It is this latter category that we refer to as ontogenic neural networks.

The other primary motivation for the project is to explore the application of neural networks, ontogenic neural networks in particular, to extrapolating, especially predicting the future behavior of a chaotic system given its past behavior. Chaos is defined as "the complicated behavior of simple deterministic equations" [Theiler, 1990].

In order to extrapolate, the network must first form an accurate model of the system's behavior. This is done by presenting exemplars of known system behavior to the neural network inputs. Network weights are adjusted to force the network outputs to match the desired outputs given by the exemplars. The trained network can then be used to produce outputs for a system state not given by the training exemplars. A properly trained network can interpolate between the training exemplars or extrapolate beyond the training exemplars.

Problem Discussion

The goal of this study is to discover how to accurately extrapolate a chaotic time series using the Cascade Correlation ontogenic neural network. Our hypothesis is that some of Cascade Correlation's network parameters affect the extrapolation accuracy. We designed experiments to find out which parameters affect training accuracy, how, and to what extent.

Cascade Correlation: An Ontogenic Neural Network

Cascade Correlation [Fahlman, 1990] begins with only input and output layers. Weights are trained using any learning algorithm that applies to such a topology. Examples are the Widrow-Hoff delta rule, the Perceptron learning algorithm, backpropagation, and the quickprop learning algorithm described in [Fahlman, 1988]. We used quickprop for this study.

When it is determined that further training is not constructive, a pool of candidate nodes for a new hidden layer are trained to capture the remaining network error. The criteria used to select the best candidate is the correlation (hence the second part of the name) between each candidate's output and the

remaining network error. When it is determined that further training of the candidates is not constructive, the candidate with the highest correlation score is installed in its own new hidden layer, with inputs from all input nodes and all previously installed hidden nodes. Because each hidden unit is in its own layer, a cascade of single-node hidden layers is formed (hence the *cascade* part of the name). Once installed, hidden nodes are never modified. Thus, they become permanent feature detectors.

After installing a new hidden node, the weights of the output nodes are retrained as described above. If the conditions for successful training have not been met, a new candidate pool is trained and the output weights are retrained. This process is repeated until the conditions for successful training have been met. The condition for successful training used in all of the experiments presented in this paper was that the error for each training exemplar must be less than one percent.

The network parameters we believe may affect extrapolation accuracy include the number of nodes in the input and output layers, the number of training exemplars, the number and nature of the nodes in the candidate pool, the criteria used to determine when further training is not constructive, and parameters which determine the rate of change of the weights over the weight space. This last set of parameters is determined by the weight-update learning algorithm chosen.

The Mackey-Glass Equation: A Chaotic Time Series

Two chaotic time series generated by the Mackey-Glass equation were used by Lapedes and Farber in their pioneering work on neural network extrapolation [Lapedes, 1987]. We chose to use one of these chaotic time series for comparability with prior works and availability of data. The Mackey-Glass equation is a nonlinear differential delay equation:

$$\frac{dx(t)}{dt} = \frac{ax(t - \tau)}{1 + x^{10}(t - \tau)} - bx(t)$$

The parameter t is the time variable, x is a function of t , and a , b , and τ (tau) are constants. We used $a = 0.2$ and $b = 0.1$. Different values of τ produce various degrees of chaos. We used $\tau = 17$, which is just beyond the onset of chaos. Solving the Mackey-Glass equation (for solution details see [Stright, 1988]) yields the time series $x(t)$. The data set used in our experiments consists of values of $x(t)$ spaced at equal time intervals Δt . A portion of this data set is shown in Figure 1.

When observing the behavior of a chaotic system, it appears that the underlying mathematics are quite complex. However, as the Mackey-Glass equation illustrates, a seemingly random, ill-behaved system can be described by a relatively simple deterministic equation. With $\tau = 17$ the Mackey-Glass time series exhibits a chaotic quasi-periodicity.

As Lapedes and Farber showed, such a non-polynomial chaotic series is difficult, if not impossible, to extrapolate using conventional polynomial extrapolation methods, but can be successfully extrapolated for a short time using neural networks. Theiler points out that "The hallmark of a chaotic system is the sensitivity of the system's individual trajectories to their initial conditions." [Theiler, 1990] For extrapolation of a chaotic time series, this means that if the initial extrapolations are not extremely accurate, the following extrapolations will stray exponentially from the correct values. Herein lies the difficulty in extrapolation when chaos is present.

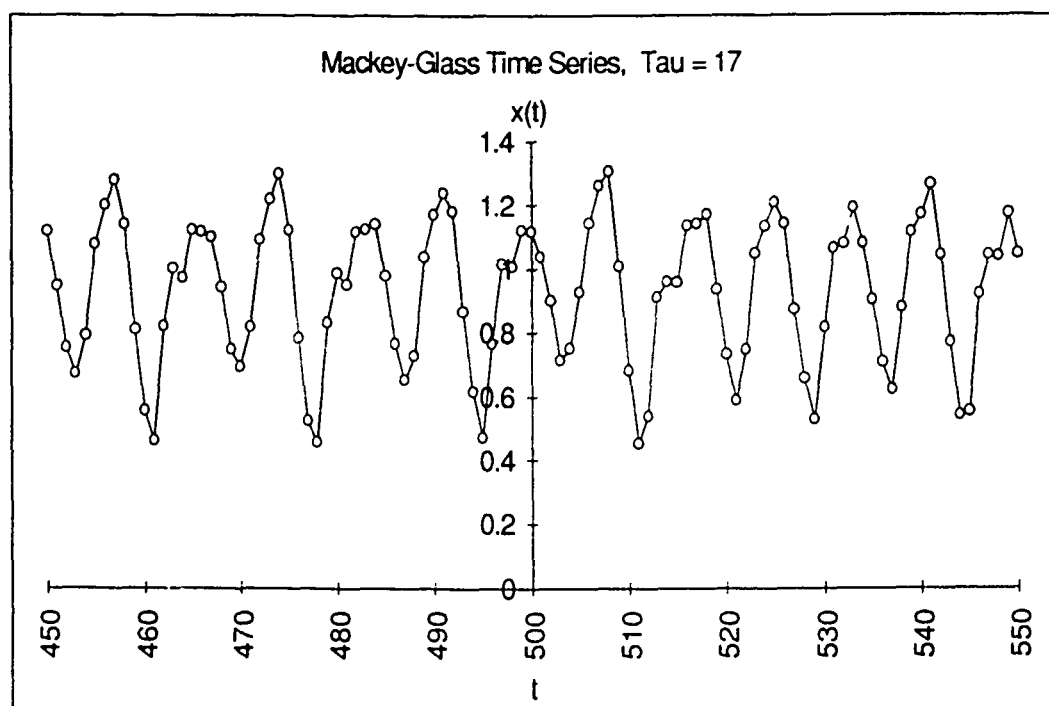


Figure 1: Time series generated using Mackey-Glass equation.

Mechanics of Extrapolation

To understand how extrapolation is accomplished using a neural network, it is first necessary to understand how an extrapolation network is trained. For an n input, m output network, a training exemplar is formed by taking $n+m$ consecutive values from the data series to be extrapolated. Starting at an arbitrary value x_i , the first n values (x_i, \dots, x_{n+i-1}) are presented to the network inputs, and the next m values ($x_{n+i}, \dots, x_{n+m+i-1}$) are compared to the network outputs for the calculation of training error. Thus a general training exemplar for an n input, m output network can be represented as $(x_i, \dots, x_{n+i-1}, x_{n+i}, \dots, x_{n+m+i-1})$. Each successive exemplar is formed by starting one value beyond the previous starting value.

We name the last known value in the data series x_k . Thus x'_{k+1} is the first value to be extrapolated. We use a prime (') to distinguish the extrapolated value from the un-primed actual value. For fair comparisons we always used $k=500$ so that the first value extrapolated was always x'_{501} (see Table 1).

After the extrapolation network is trained, the first extrapolation x'_{k+1} is made by presenting the last n values in the data series (x_{k-n+1}, \dots, x_k) to the network inputs. Thus for x'_{k+1} , all inputs are known data values. The second extrapolation x'_{k+2} is made by presenting ($x_{k-n+2}, \dots, x'_{k+1}$) to the network inputs. The successive input exemplars are likewise shifted by one. For x'_{k+2} one input is an extrapolated value, for x'_{k+3} two inputs are extrapolated values, etc. Eventually all extrapolations are made based upon previously extrapolated data.

<u>Values Presented to Network Inputs</u>	<u>Values Compared with Network Outputs</u>
:	:
$x_{494}, x_{495}, x_{496}, x_{497}$	x_{498}
$x_{495}, x_{496}, x_{497}, x_{498}$	x_{499}
$x_{496}, x_{497}, x_{498}, x_{499}$	x_{500} (Last training exemplar)
$x_{497}, x_{498}, x_{499}, x_{500}$	x'_{501} (First extrapolation)
$x_{498}, x_{499}, x_{500}, x'_{501}$	x'_{502} (Second extrapolation obtained using first extrapolation)
:	:

Table 1: Examples of training exemplars and extrapolations for $n = 4$, $m = 1$, $k = 500$.

Evaluation of extrapolation accuracy

After the first few experiments were run and examined, it became clear that determining which extrapolations are best is not a simple problem. Some extrapolations were very good for a short period only; others were not very close but did not stray far. Which of these is best may depend on the application. Therefore, we need ways of evaluating extrapolation accuracy which will allow us to select the best network for a particular application.

Plotting the extrapolation curve vs. the actual curve

One common method for examining extrapolation results is to plot the extrapolation curve along with the correct curve. For a quasi-periodic function such as the Mackey-Glass equation, it seems logical to describe this plot in terms of four measurements of periodic functions: offset, amplitude, phase, and frequency.

Offset error was only noticed in our earliest experiments when the training criteria were not very strict (e.g., only trained to ten percent error for each exemplar). The extrapolation curve followed the actual curve, but at a different height. We call this problem offset error because it resembles the adjustment of the dc offset of a waveform as viewed on an oscilloscope. To get rid of offset error, train further or retrain with stricter training criteria.

Amplitude errors almost always accompanied offset errors. We encountered three types of amplitude errors: permanent, temporary, and growing (exploding).

As with offset errors, permanent errors in amplitude only occurred with slack training criteria. Interestingly enough, when these errors occurred, the shape of the extrapolation curve otherwise matched the shape of the actual curve; the network had begun to learn the underlying mathematical model. To get rid of permanent amplitude error, train further or retrain with stricter training criteria.

Sometimes the amplitude of the extrapolation curve would get temporarily larger or smaller than that of the actual curve. This may be due to the fact that the chaotic data of the Mackey-Glass equation has temporary changes in amplitude. If this is true, then the network learned of these changes and was simply predicting these changes in the wrong places.

Another temporary amplitude error was observed for which we have no explanation. We call it *tunneling* because of its shape. Sometimes tunneling appeared as a flat or nearly flat portion on the extrapolation curve (Figure 2). Other times tunneling appeared as small, quick oscillations (Figure 3).

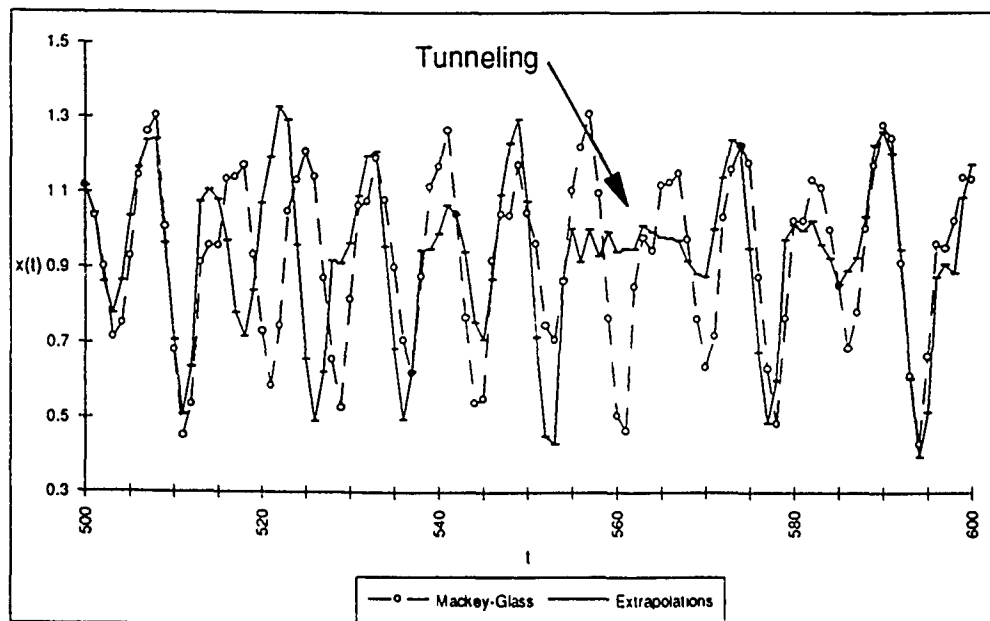


Figure 2: Tunneling example; nearly flat.

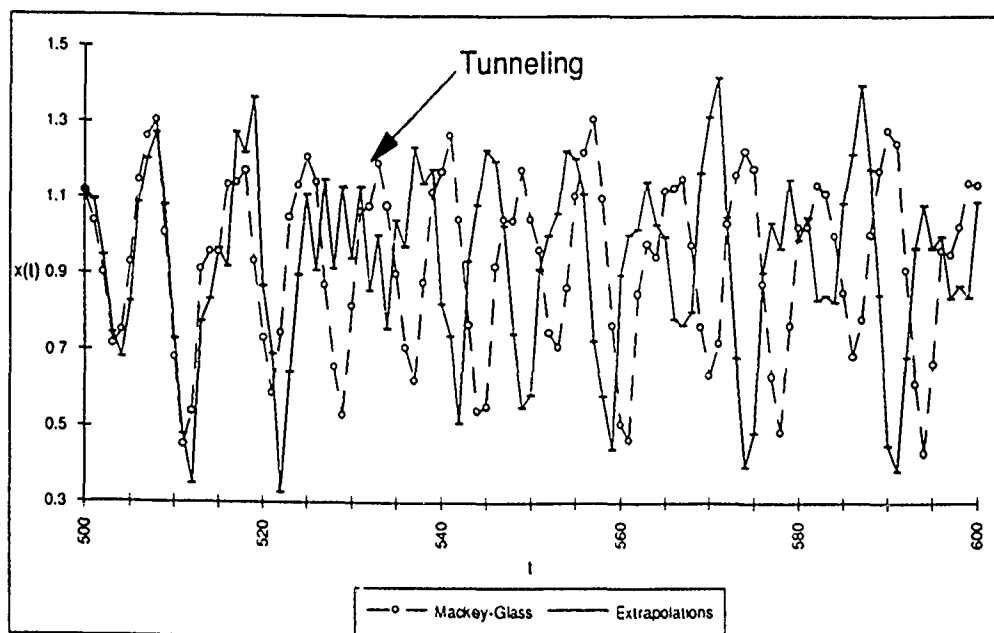


Figure 3: Tunneling example; small, quick oscillations.

The worst possible extrapolation error is the growing amplitude error. It comes in three varieties: the extrapolation increases exponentially to positive infinity, the extrapolation decreases exponentially to negative infinity, or the extrapolation is a quasi-sinusoid with exponentially growing amplitude. We refer to this type of error as an exploding extrapolation.

Most of the phase errors we found were lagging phase errors: the extrapolation curve lagged the actual curve. In a small number of cases the phase error was leading.

Some phase errors are due to frequency errors. This produces an interesting result. If the frequency of the extrapolation curve is lower than that of the actual curve, the extrapolation lags at first with growing phase error. After being totally out of phase the extrapolation will lead. Eventually the extrapolation will temporarily get back on track.

Plotting the extrapolation error curve

Another useful plot describing individual extrapolation errors is the *extrapolation error curve*, a plot of the extrapolated values minus actual values. When the extrapolation error curve is positive, the extrapolation is greater than the actual; when the extrapolation error curve is negative, the extrapolation is less than the actual. This plot gives a truer, more informative, and more easily read graph of individual extrapolation errors than does the plot of the extrapolation along with the actual, which can be misleading.

An example of how the extrapolation vs. actual curve can be misleading is shown in Figures 4 and 5. Notice that though the extrapolation curve appears to closely trace the actual curve for the first 45 extrapolations, individual extrapolation errors are large. Compare with the extrapolation error curve, which quickly identifies the poor extrapolation.

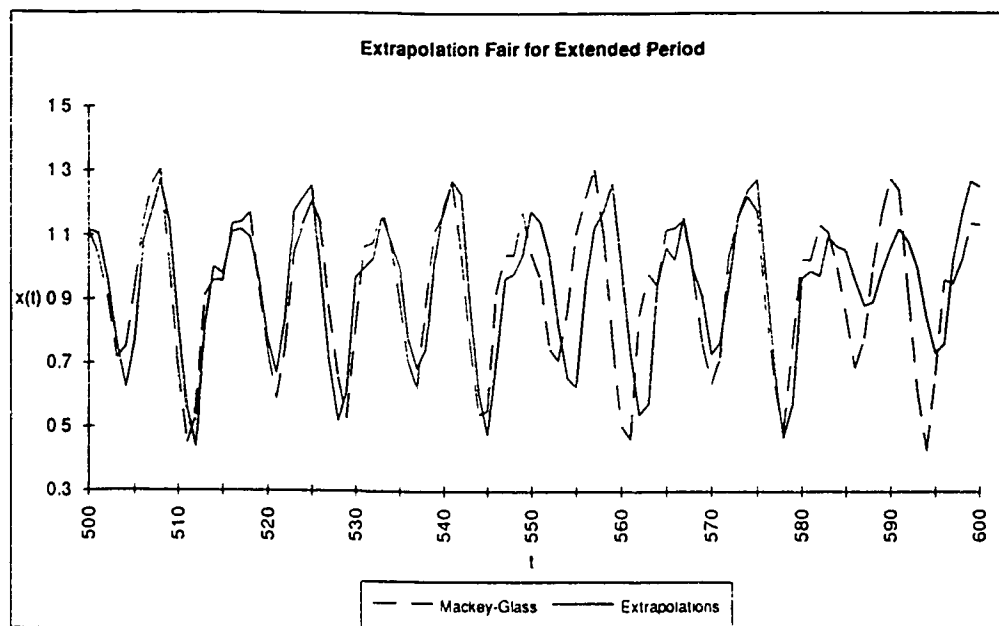


Figure 4: Misleading representation of extrapolation errors.

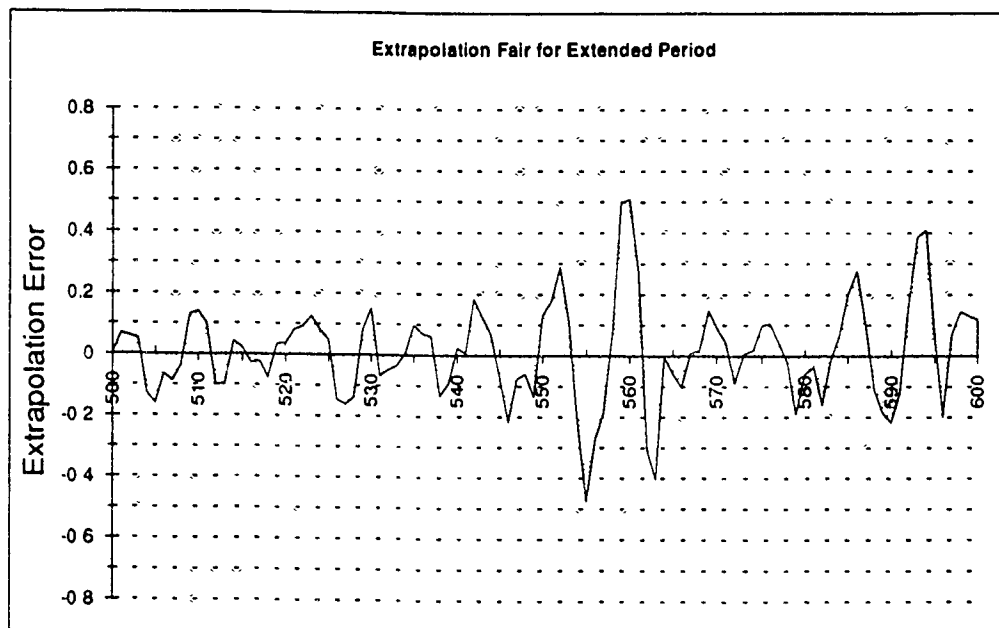


Figure 5: Truer representation of extrapolation errors.

Figures 4 and 5 also illustrate a case in which extrapolations are never very close, but neither do they stray far. Contrast this case with the case illustrated in Figures 6 and 7 in which extrapolations are very good (error less than 10%), but only for a short period (about 35 extrapolations).

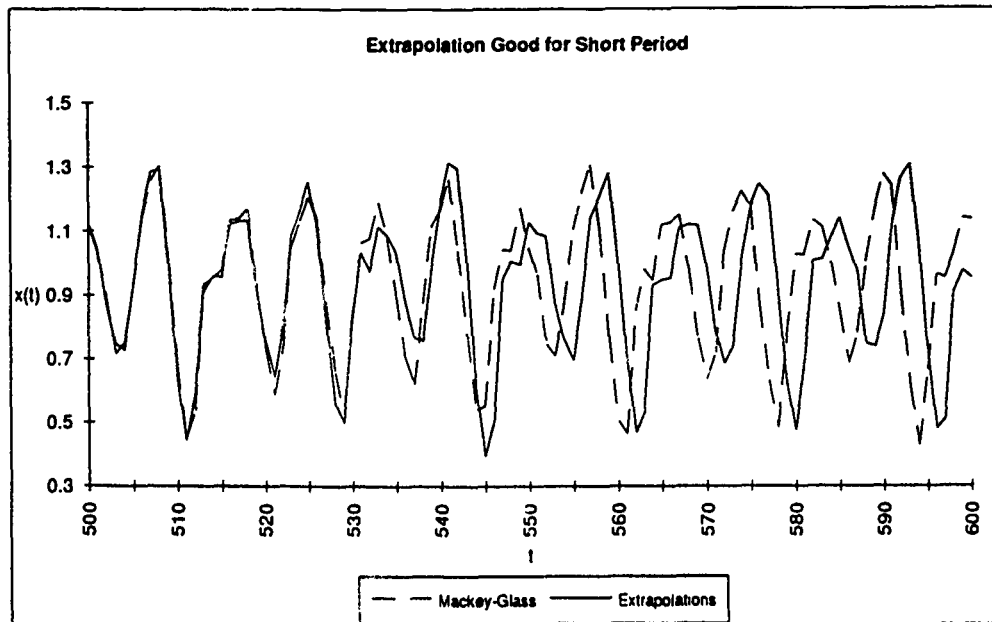


Figure 6: Extrapolation good for a short period.

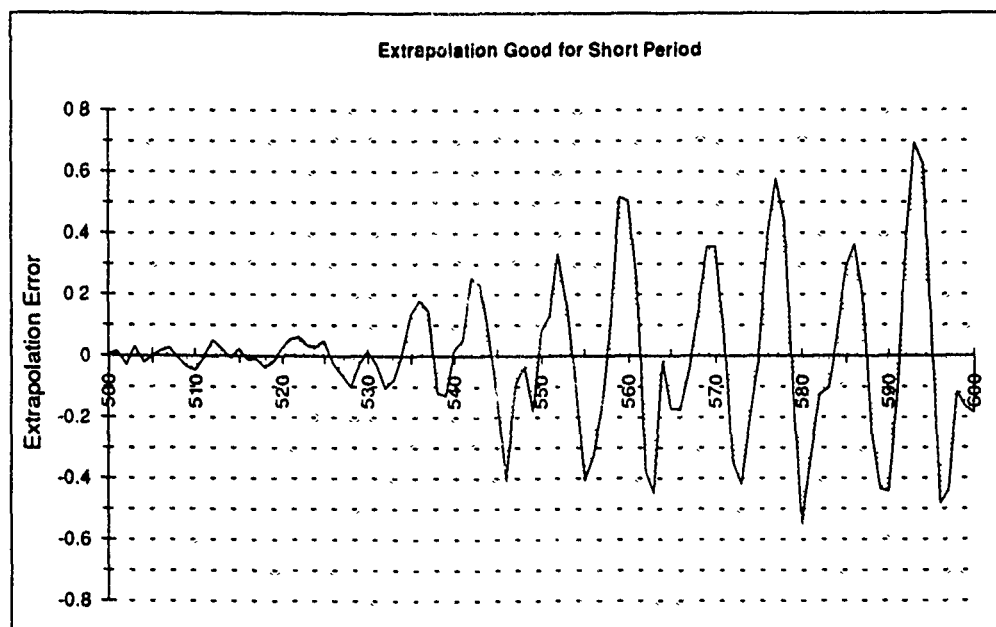


Figure 7: Extrapolation good for a short period.

Calculating the RMS error for a given range of extrapolations

The RMS error is a popular error measure. It is effective for comparing extrapolations of different networks over a particular range. It offers the advantage of a single value for measuring extrapolation accuracy, but at the expense of information. This requires that one use caution when using the RMS error for comparing different networks. Whereas one network may have the lowest RMS error over one range, a different network may have the lowest RMS error over another range. The desired criteria of extrapolation accuracy will help determine the proper range to use for the RMS error. Caution must still be used, for networks with similar RMS errors over one range may rank differently if the range is changed even slightly.

Plotting the cumulative extrapolation error curve

The *cumulative extrapolation error curve* overcomes this disadvantage of RMS error calculations while allowing easier comparison of different networks than the other plots discussed in this paper. At each point the cumulative extrapolation error curve displays the total of the absolute values of all extrapolation errors up to and including that particular point. The absolute value is used so that high and low extrapolations do not cancel each other's effect on this error curve. This plot will help identify the points at which changes in rank (by extrapolation accuracy) occur. Notice in Figure 8 that Net 2 has better extrapolation accuracy only if the number of extrapolations needed by an application is between 16 and 32.

The cumulative extrapolation error curve also provides a good check of the validity of RMS error comparisons. In Figure 9 notice that Net 3 has the smallest error for 100 extrapolations. If only the RMS error was investigated, Net 3 would be incorrectly chosen as the best network. Actually, Net 3 provided the poorest extrapolation accuracy, while Nets 1 and 2 were quite good for a short period.

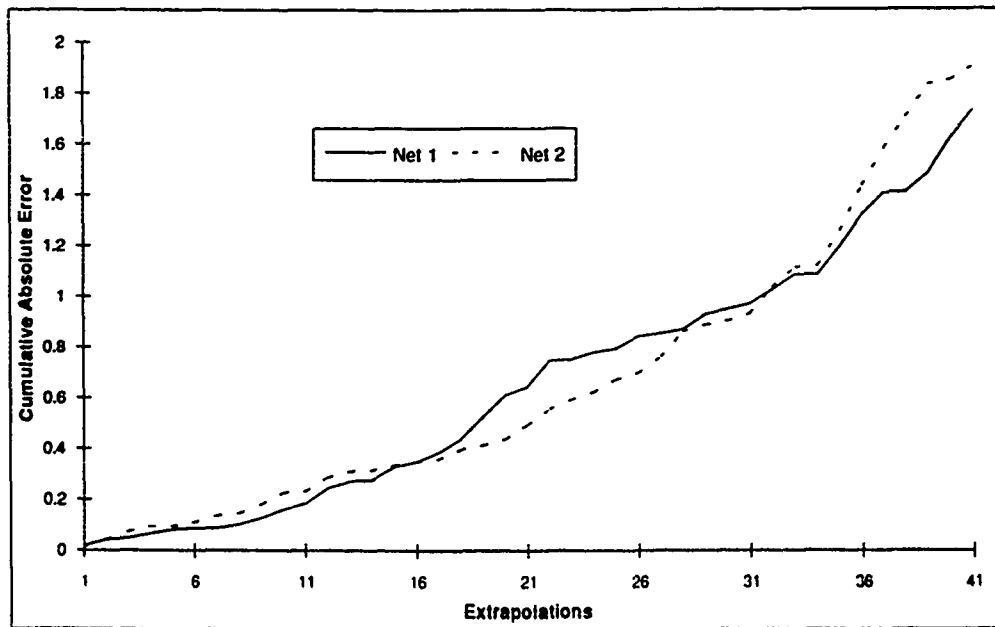


Figure 8: Cumulative extrapolation error curves show best extrapolations for any range.

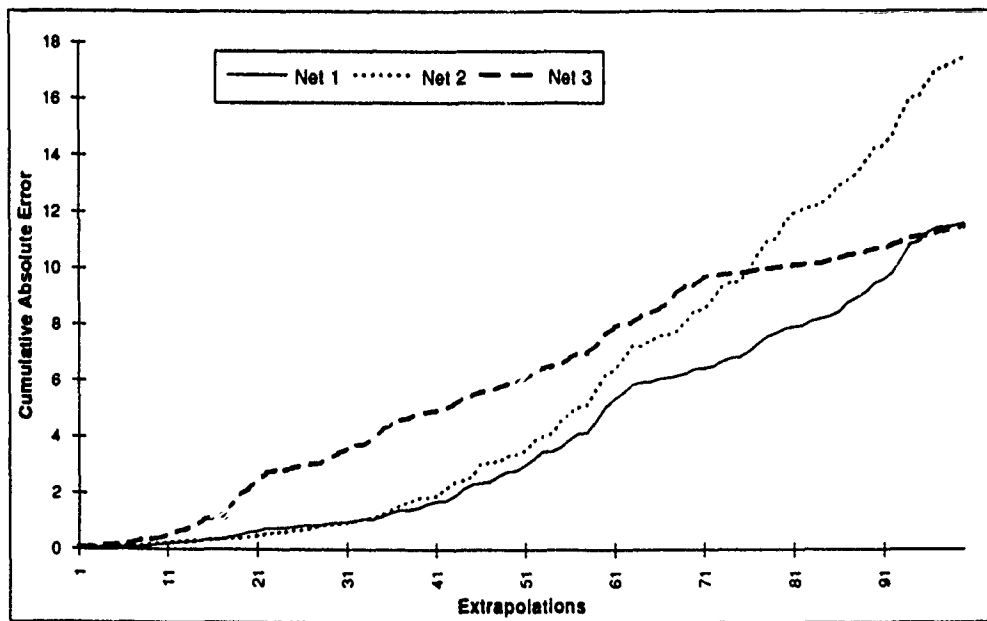


Figure 9: Cumulative extrapolation error curves show where RMS error is misleading.

Experiments

We feel that most of the keys to accurate extrapolation lie in understanding how network parameters affect extrapolation accuracy. Early experiments yielded some insight into how the nature of candidate nodes affects extrapolation accuracy. Two experiments specifically designed to investigate other network parameters have been completed thus far. One investigates how the numbers of inputs and outputs affect extrapolation accuracy. The other investigates how the number of training exemplars affects extrapolation accuracy.

Activation functions

The source code we used for the Cascade Correlation algorithm was provided by Dr. Scott Fahlman and Scott Crowder of Carnegie Mellon University. It included the Gaussian, logistic sigmoid, and symmetric logistic sigmoid activation functions. We added several activation functions to the Cascade Correlation algorithm (see Tables 2 through 4). We then modified the algorithm to allow more than one activation function in the candidate pool. This meant that nodes with different activation functions could compete simultaneously. Cascade Correlation automatically selects the best candidate, and thus helped us determine which activation functions are best for our data.

When given a choice among the activation functions in Tables 2 through 4, Cascade Correlation chose sines and cosines almost exclusively and in nearly equal quantities. Of networks trained with only one activation function, sines and cosines outperformed all others in extrapolating. This is not surprising considering the quasi-sinusoidal data set (refer to Figure 1).

Of the Gaussian types, the negative Gaussian performed best on our data set. The Gaussian also performed satisfactorily. Compare the Gaussian with the hyperbolic secant. The latter approaches zero more slowly. Surprisingly, there was a marked difference between the two in training; for our data set, the hyperbolic secant was the worst of all activation functions we tried.

For our data set, sigmoids in the range $[-1, +1]$ had more success in training than those of other ranges. The sigmoid most often chosen by Cascade Correlation was the double symmetric logistic sigmoid. The hyperbolic tangent also performed satisfactorily. The former has a smoother transition.

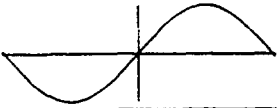

Activation Function	Formula	Graph	Range
Sine	$f(x) = \sin(x)$		$[-1,+1]$
Cosine	$f(x) = \cos(x)$		$[-1,+1]$

Table 2: Sinusoidal hidden activation functions.


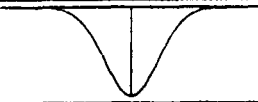
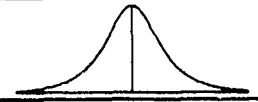
Activation Function	Formula	Graph	Range
Gaussian	$f(x) = e^{-(x^2/2)}$		$[0,1]$
Negative Gaussian	$f(x) = -e^{-(x^2/2)}$		$[-1,0]$
Hyperbolic Secant	$f(x) = \text{sech}(x)$		$[0,1]$

Table 3: Gaussian and similar hidden activation functions.


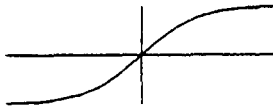
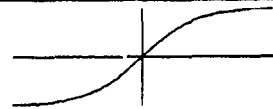
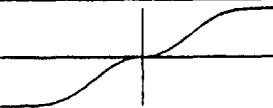
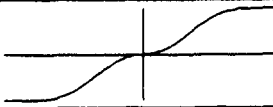
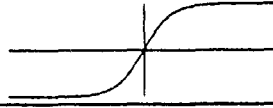
Activation Function	Formula	Graph	Range
Logistic Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$		$[0,1]$
Symmetric Logistic Sigmoid	$f(x) = \frac{1}{1+e^{-x}} - 0.5$		$[-0.5,+0.5]$
Double Symmetric Logistic Sigmoid	$f(x) = \frac{2}{1+e^{-x}} - 1$		$[-1,+1]$
Symmetric Logistic Sigmoid with Square *	$f(x) = \frac{1}{1+e^{-x^2}} - 0.5$		$[-0.5,+0.5]$
Double Symmetric Logistic Sigmoid with Square *	$f(x) = \frac{2}{1+e^{-x^2}} - 1$		$[-1,+1]$
Hyperbolic Tangent	$f(x) = \tanh(x)$		$[-1,+1]$

Table 4: Sigmoidal hidden activation functions.

*Invented by Dale Nelson and David Ensley

The author developed the equations for the sigmoids with square terms based upon a graph proposed by Dale Nelson. Nelson believes the flat spot around the origin will help eliminate noise; small weight perturbations during training would be dampened. Though these functions' performance on our data set was only fair, Nelson plans to try them on different data sets to test his hypothesis.

Numbers of inputs and outputs

Takens' theorem tells us that the number of inputs n to an extrapolating neural network should be greater than d and less than $2d+1$, where d is the fractal dimension (a measure of the degree of chaos [Gearhart, 1991]) of the data set [Lapedes, 1987]:

$$d \leq n \leq 2d + 1$$

The fractal dimension of the time series generated by the Mackey-Glass equation with $\tau = 17$ is approximately 2.1 [Lapedes, 1987 & Gearhart, 1991]. Thus Takens' theorem tells us that we need 3, 4, or 5 inputs for accurate extrapolation.

Following the example of [Lapedes, 1987] and [Stright, 1988], all of our early experiments used 4 inputs and 1 output. To investigate the effect of the number of inputs and outputs on extrapolation accuracy, we chose one network whose extrapolations were worse than average in accuracy but not exploding. Keeping all other parameters unchanged (including the seed for the random number generator), we retrained this network with every combination of 2 through 10 inputs and 1 through 3 outputs. (Unfortunately, time restraints allowed us only one run for each case.) The two-output case forces the network to learn the first derivative (slope) of the actual curve along with future points. The three-output case forces the network to additionally learn the second derivative (curvature). The second and third outputs are used for training purposes only and hence are ignored during extrapolation.

The resulting RMS errors for the first 100 extrapolations are shown in Figure 10. The absence of a column indicates that the extrapolation exploded with an RMS error greater than 2. For each number of outputs, data from the best results appear in Table 5. The best results were all in the range of inputs predicted by Takens' theorem. More precisely, the best range of inputs for this experiment was between d and $2d$. With only one run for each combination of inputs and outputs, we cannot say conclusively that Takens' range was too broad.

Using more outputs seems to produce fewer exploding errors at the cost of overall extrapolation accuracy. Again, we need more data to be sure.

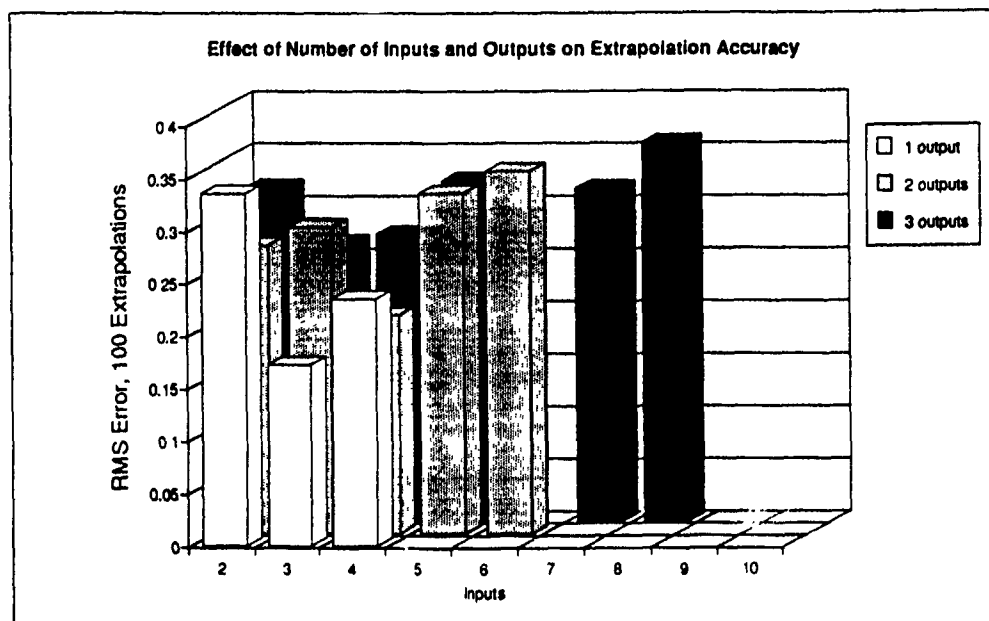


Figure 10: Inputs/outputs experiment results.

# Inputs	# Outputs	# Training Epochs	# Hidden Nodes	RMS
3	1	9626	42	0.172
4	2	15021	62	0.208
3	3	13549	69	0.265

Table 5: Best results from inputs/outputs experiment for 1, 2, and 3 outputs.

Number of training exemplars

To investigate the effect of the number of training exemplars on extrapolation accuracy, we varied the number of exemplars from 50 to 450 in steps of 50 and ran 5 cases of each. Except for the number of training exemplars, each run differed only in the initial random weights. All of the networks in this experiment used 4 inputs and 1 output. All training sets ended at x_{500} so that x_{501} was always the first point extrapolated.

The best 3 runs (based on lowest RMS errors for the first 100 extrapolations) in each category are shown in Figure 11. The absence of a column indicates a very poor extrapolation, most of which exploded. Notice the trend toward a minimum extrapolation error around 300 exemplars. It seems that for small numbers of exemplars, the network does not receive enough information about the underlying function to consistently model it accurately, while large numbers of exemplars keep the network from generalizing. We suspect that, as is true with the number of inputs, the best number of training exemplars is closely related to the fractal dimension of the data.

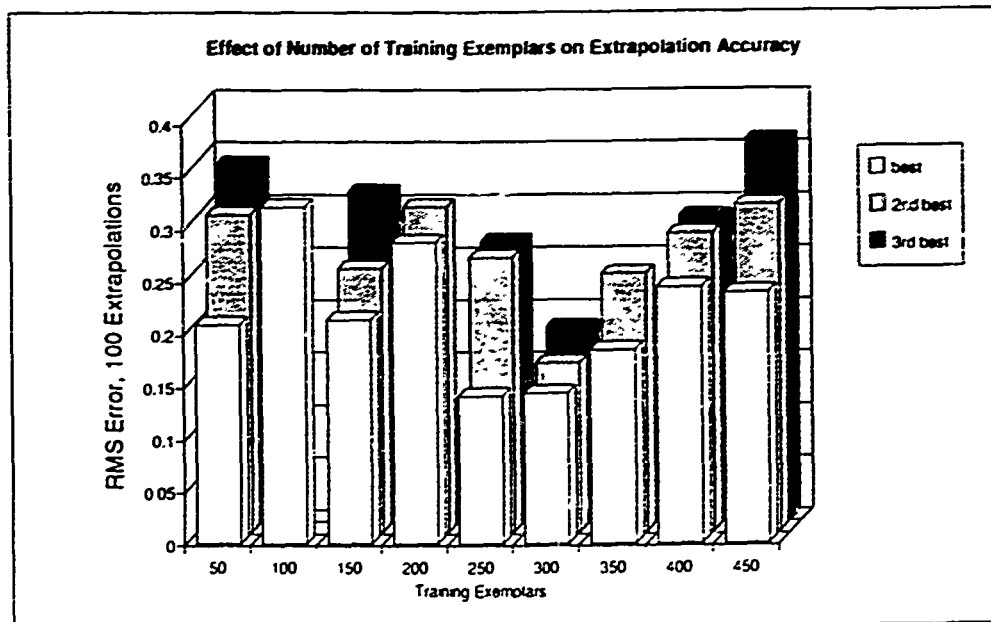


Figure 11: Exemplars experiment results.

# Exemplars	# Training Epochs	# Hidden Nodes	RMS
250	8434	34	0.141
300	9032	36	0.144
300	10275	40	0.161
300	11232	46	0.183
350	12460	49	0.184

Table 6: Five best results from exemplars experiment.

Conclusions

One of the major concerns in developing a neural network for extrapolation is the possibility of exploding extrapolation errors. Several trained networks that perform well on the training exemplars can have very different extrapolations. A few well-trained networks may have exploding extrapolation errors. At this time we do not know how to tell in advance if a well-trained network's extrapolations will explode or not. Neither can we tell in advance how accurate the extrapolations will be if they do not explode. For most applications the existence of exploding extrapolations can be easily tested; this is not the case for applications requiring real-time learning. More research is needed to determine how to prevent both inaccurate and exploding extrapolations. As mentioned earlier, we have found ways of preventing two types of extrapolation error, namely offset and permanent amplitude errors.

Cascade Correlation can help select the best hidden layer activation functions for a particular data set by using a mixture of activation functions in the candidate pool. Proper selection of the activation function, the number of training exemplars, the numbers of inputs and outputs, and possibly other network parameters provides greater extrapolation accuracy. We are beginning to discover how to properly select these parameters.

We have shown that when judging networks on extrapolation accuracy, plotting the extrapolation curve vs. the actual curve and using the RMS error over extrapolations can both be misleading. The extrapolation error curve and cumulative extrapolation error curve, respectively, are useful for checking these measurements. In addition, the extrapolation error curve is much easier to read, and the cumulative extrapolation error curve quickly identifies the best networks for the extrapolation ranges of interest.

Acknowledgments

The author would like to thank the Air Force Office of Scientific Research for providing this excellent summer research opportunity and Research and Development Laboratories and Milton H. Danishek for their part in administering this program. Dr. J. David Irwin and Dr. Charles A. Gross of Auburn University helped make this effort possible. Thanks to Edward Gliatti and staff at the Information Processing Technology Branch of the Avionics Laboratory for making my stay very pleasant. Thanks to Mason Friar for providing an excellent computing platform for this research. Dr.

Louis Tamburino, Dr. Tom Gearhart, Dennis Buck, and Eileen Lantz provided assistance, resources, and insight. Thanks to Dr. Scott Fahlman and Scott Crowder of Carnegie Mellon University for the source code for Cascade Correlation and for their assistance and recommendations. Special thanks to Dale Nelson, my effort focal point and Chief of the Advanced Systems Research Group, for sharing his insight, knowledge, guidance, experience, and much time.

References

- [Alpaydin, 1991] Alpaydin, E., "GAL: Networks That Grow When They Learn and Shrink When They Forget," Technical Report TR 91-032, International Computer Science Institute, Berkeley, CA, May 1991.
- [Barnhart, 1970] Barnhart, C. L., Editor in Chief, *The American College Dictionary*, Random House, New York, 1970.
- [Fahlman, 1988] Fahlman, S. E., "An Empirical Study of Learning Speed in Back-Propagation Networks," Technical Report CMU-CS-88-162, Carnegie Mellon University, Pittsburgh, PA, September 1988.
- [Fahlman, 1990] Fahlman, S. E. and C. Lebiere, "The Cascade-Correlation Learning Architecture," Technical Report CMU-CS-90-100, Carnegie Mellon University, Pittsburgh, PA, February 1990.
- [Gearhart, 1991] Gearhart, T. K., "Correlation Dimension of Chaotic Attractors," Technical Report, Avionics Directorate, Wright-Patterson Air Force Base, Dayton, OH, July 1991.
- [Lapedes, 1987] Lapedes, A. and R. Farber, "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling," Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM, June 1987.
- [Stright, 1988] Stright, J. R., "A Neural Network Implementation of Chaotic Time Series Prediction," Thesis, Air Force Institute of Technology, December 1988.
- [Theiler, 1990] Theiler, J., "Estimating the Fractal Dimension of Chaotic Time Series," *The Lincoln Laboratory Journal*, vol. 3, no. 1, pp. 63-85 (1990).

COMPUTER AIDED ANALYSIS OF LPI SIGNAL DETECTABILITY

Scott P. Francis

Graduate Student, University of Kansas

Abstract

A low probability of intercept signal detectability analysis program is described which calculates the detectability of certain spread spectrum signals by radiometric detection systems and feature detectors. LPI/SDA is a PC-based computer aided analysis system which expresses the detectability of a signal in terms of the signal carrier power to one-sided noise power spectral density C/N_{01} required at the input to a radiometer or feature detector to achieve user-specified detection and false alarm probabilities. LPI/SDA also has the ability to calculate LPI system Quality Factors, which describe the detectability of a signal separately in terms of scenario dependent and scenario independent factors. Sample detectability calculations are illustrated, and performance curves are provided. This program was initially developed at the Telecommunication and Information Sciences Laboratory (TISL) at the University of Kansas, and was modified and enhanced at the Avionics Laboratory at Wright-Patterson AFB during the 1991 AFOSR Summer Research Program.

1 Introduction

Military RF communication systems must necessarily provide a high level of security against the exploitation of transmitted information by an unintended listener. This exploitation could be as simple as detecting the presence and location of a communications platform, or as complex as detecting the presence of and extracting the information contained in a transmitted signal. Nicholson [5] describes four sequential operations that exploitation systems attempt to perform:

1. Cover the signal, that is, have a receiver tuned to some or all of the frequency intervals being occupied by the signal when the signal is actually being transmitted.
2. Detect the signal, that is, make a decision about whether the power in the intercept bandwidth is a signal plus noise and interference or just noise and interference.
3. Intercept the signal, that is, extract features of the signal to determine if it is a Signal of Interest (SOI) or a Signal Not of Interest (SNOI).
4. Exploit the signal, that is, extract additional signal features as necessary and then demodulate the baseband signal to generate a stream of binary digits.

The probability of exploitation $Pr(E)$ is given as

$$Pr(E) = Pr(E|I) Pr(I|D) Pr(D|C) Pr(C)$$

where $Pr(E|I)$ is the probability of exploitation given interception, $Pr(I|D)$ is the probability of interception given detection, $Pr(D|C)$ is the probability of detection given signal cover, and $Pr(C)$ is the probability of signal cover. Everything that an unintended listener could conceivably want to do with a signal depends critically on having the ability to cover and detect the presence of the signal. Any subsequent actions are dependent upon this signal detection.

Military communication system designers have traditionally employed spread spectrum waveforms to achieve a particular level of transmitted signal covertness. These spread spectrum signals, in addition to permitting the use of code division multiple access (CDMA) for efficient bandwidth utilization, also incorporate significant anti-jam (AJ) and low probability of intercept (LPI) characteristics due to their low-level radiated power densities. The term LPI is used here as it is in much of the literature (e.g., [4]), although LPI signals are perhaps better described as low probability of detection (LPD) signals. LPI will be used in this document to describe signals which are difficult for an unintended receiver to detect.

The communications receiver in an LPI communication system possesses knowledge of the code which was used at the transmitter to spread the signal, and thus can despread the received signal by remixing it coherently with the code. This despreading operation allows the receiver to filter out a large portion of the noise power present within the spread bandwidth at the receiver front-end. An unintended receiver does not typically have knowledge of this spreading code and must make signal present decisions based solely on the received energy in some frequency band over some period of time. Further, because the unintended receiver lacks the ability to despread the signal, it is unable to filter out any of the noise power within the spread bandwidth. Receivers which make binary signal present decisions based on energy detection are called radiometric systems (radiometers), and represent the most common detection threat to LPI signals.

Another type of common detection threat is the feature detector. A feature detector processes the received signal in such a way that spectral lines appear in the spectrum of the output waveform which were not apparent in the spectrum of the received signal. These spectral lines, in addition to indicating signal presence, may also contain information concerning the received signal's center frequency, chip rate, and hop rate, and thus may assist the interceptor in identifying the signal.

The inherent vulnerability of an LPI (spread spectrum) signal to detection by a particular radiometric system or feature detector can be quantified in terms of the required carrier signal power to one-sided noise power spectral density ratio C/N_{01} required at the front end of the interceptor to achieve a specified probability of detection P_D and probability of false alarm P_{FA} performance level [1, 11]. The LPI

communication system designer uses this detectability information to select the spread spectrum modulation type and parameters to yield a signal which is minimally detectable by the most likely detection threat, in this case a particular type of radiometer system or feature detector.

Many researchers have developed analytical models which map the detector performance probabilities to the required front-end C/N_{01} , e.g. [1]. LPI/SDA is a PC-based computer aided analysis system which automates the use of these models. The LPI communication system designer interacts with LPI/SDA's user-friendly hierarchical interface to describe a signal and detector system, and quickly determines the required C/N_{01} . Signal and detector parameters can be easily modified to evaluate the effects of these changes on the required C/N_{01} . LPI/SDA also has the ability to calculate five different LPI system Quality Factors. These Quality Factors, which describe the detectability of a signal separately in terms of scenario dependent and scenario independent factors [4], are described further in section 5.

2 LPI Signal Models

LPI/SDA models three standard and four hybrid types of spread spectrum signals. They are:

- Direct Sequence (DS) – also known as pseudonoise
- Frequency Hopped (FH)
- Time Hopped (TH)
- Frequency Hopped/Direct Sequence (FH/DS)
- Time Hopped/Direct Sequence (TH/DS)
- Frequency Hopped/Time Hopped (FH/TH)
- Frequency Hopped/Time Hopped/Direct Sequence (FH/TH/DS)

A direct sequence signal is one which is obtained by mixing a modulated signal directly with a pseudo-random, typically binary, code. The spread spectrum bandwidth of the DS signal is much greater than the bandwidth of the original signal, which typically has a time-bandwidth product of approximately unity, and thus the time-bandwidth product of the DS signal is much greater than one. By utilizing a large time-bandwidth product the signal is effectively "buried" in the noise, making its presence difficult to detect. The intended communications receiver has the ability to perform coherent detection of the DS signal and can "lift" the signal out of the noise floor. The unintended receiver lacks this ability and must make signal

present decisions based on received signal and noise energy or some other feature of the received signal.

A frequency hopped signal contains unity time-bandwidth pulses which are pseudo-randomly placed throughout the spread spectrum bandwidth during the message transmission. The duty cycle of frequency hopped signals is one, i.e., pulses are transmitted continuously for the entire message duration. The intended communications receiver has knowledge of the pulse positions in frequency as a function of time, and can coherently mix these pulses down to a single IF band. The unintended receiver does not have knowledge of the pulse positions and must either place a radiometer at every frequency hop band (there may be as many as 10,000 bands or more) or place radiometers at some subset of these bands to yield a more realizable detector (some feature detectors, however, can operate over the entire wideband bandwidth).

Time hopped signals also contain unity time-bandwidth pulses. However, instead of randomly placing these pulses in frequency, they are random placed in time, and always occupy the same frequency band. The duty cycle of these signals is something less than $1/2$. The intended receiver retains a performance edge over the unintended receiver by having knowledge of the pulse positions in time.

The frequency hopped/direct sequence hybrid signal is similar to the frequency hopped signal, except that pulses with large time-bandwidth products are used instead of unity time-bandwidth pulses. The use of this type of signal yields the benefits of both FH and DS signals, i.e., the pulses avoid radiometers by being randomly placed in the spread spectrum bandwidth and the pulses are buried in the noise.

The time hopped/direct sequence hybrid signal, like the FH/TH signal, also utilizes large time-bandwidth pulses. The benefits of this signal are analogous to those of the FH/DS signal.

Frequency hopped/time hopped hybrid signals consist of unity time-bandwidth pulses which are pseudo-randomly placed in both frequency and time. Frequency hopped/time hopped/direct sequence signals are similar except that they consist of large time-bandwidth pulses pseudo-randomly placed in frequency and time.

From an energy detection standpoint, these spread spectrum signals can be described in terms of relatively few parameters. These parameters are listed below and illustrated in the time-frequency plane shown in Figure 1.

- T_1 - message duration (sec)
- W_1 - spread spectrum bandwidth (Hz)
- T_2 - pulse duration (sec)
- b_2 - number of pulses in T_1
- T_{p2} - time hop interval (sec)

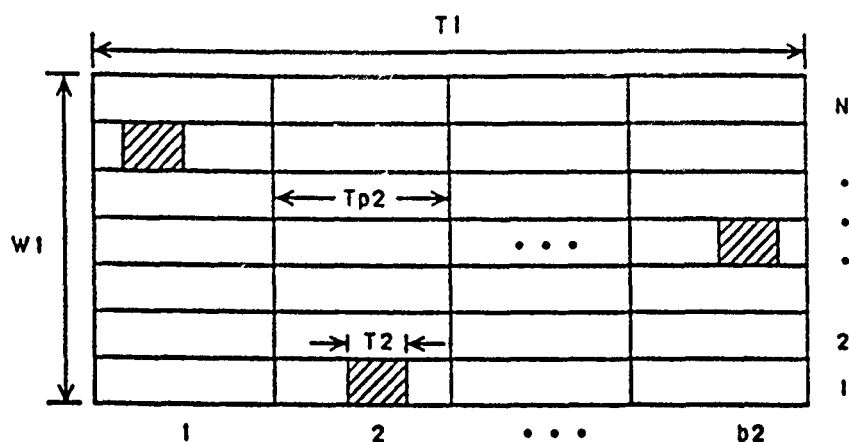


Figure 1: LPI/SDA signal parameter notation. Adapted from [1].

- N - number of frequency hop bands in W_1

To describe a DS signal, for instance, one would set $N = 1$, $b_2 = 1$, and $T_2 = T_1$. T_1 and W_1 would be set to the desired message time and spread bandwidth, respectively.

3 Radiometer Theory

The heart of all of the radiometric systems which LPI/SDA models is the wideband radiometer (also known as an energy detector or total power radiometer). This system, illustrated in Figure 2, filters a portion of the RF spectrum, squares this filtered signal to obtain signal power, and integrates from $t - T$ to t to yield signal energy (typically this integration is implemented as integrate and dump rather than continuous integration). This signal energy is then compared to a threshold and, if the threshold is exceeded, a signal is declared present; otherwise no signal is declared present. Assuming ideal signals and filters, the wideband radiometer can equivalently be described as a system which observes a rectangular time-frequency "cell" with bandwidth equal to the bandpass filter bandwidth W and time interval equal to the integration time T . It measures the total signal plus noise energy received in that cell, and compares this received energy to a threshold. A signal is declared present if the cell energy exceeds the threshold.

The performance of the wideband radiometer is characterized in terms of two probabilities: the probability of detection P_D and the probability of false alarm P_{FA} . The probability of detection is defined as the probability that a signal coincides with at least a portion of the radiometer bandwidth, during at least some of the radiometer

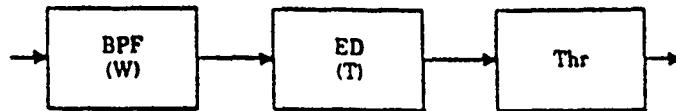


Figure 2: Wideband radiometer system (bandpass filter, energy detector, and threshold). [1]

integration time, and the radiometer declares a signal present. The probability of false alarm is defined as the probability that no signal coincides with the radiometer bandwidth during the integration time, and the radiometer nevertheless declares a signal present.

For signals which completely occupy a single time-bandwidth cell (e.g., DS signals or a single pulse of a hopped signal) and are embedded in additive white Gaussian noise, wideband radiometers represent essentially the best performing detection systems which can be constructed [10]. If a signal is pulsed in time and/or frequency, the interceptor may be able to improve his detection performance significantly if he has knowledge of the pulse positions in both time and frequency and exploits this knowledge by using an appropriate radiometer system [1]. These radiometer systems consist of one or more wideband radiometers, each with a bandwidth and integration time matched to the time-bandwidth dimensions of a pulse. The binary signal present/not present decisions that each of these radiometers makes are processed in some manner to minimize the overall false alarm probability while maintaining a high overall detection probability. These systems are described further in section 4. Note, however, that because these systems consist of conglomerations of wideband radiometers, the system performance probabilities P_D and P_{FA} can be mapped to single cell probabilities Q_D and Q_{FA} which represent the performance levels required at each of the wideband radiometers. Thus, the wideband radiometer model can be used for all types of radiometer systems once the performance probabilities have been modified appropriately.

LPI/SDA's model of the wideband radiometer is based on the following. When only Gaussian noise is present at the input to a wideband radiometer, the output follows a chi-square probability density function (PDF) with $2TW$ degrees of freedom [11]. If we let v represent the decision statistic at the output of the radiometer, then the radiometer output PDF given that noise only is present can be described

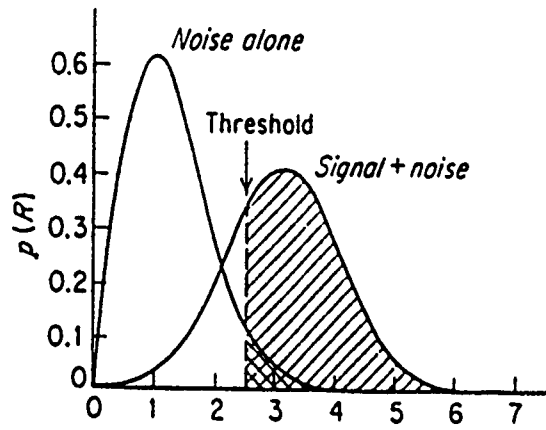


Figure 3: Example conditional radiometer output PDFs. Adapted from [6].

mathematically as [9]

$$f(v|\text{noise}) = \frac{1}{\sigma^{2TW}\Gamma(TW)} v^{TW-1} e^{-v/2\sigma^2}, \quad 0 < v < \infty \quad (1)$$

$$= 0 \quad \text{elsewhere} \quad (2)$$

where σ is the standard deviation of the Gaussian noise process and the gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \quad \alpha > 0 \quad (3)$$

With signal present, the output has a noncentral chi-square PDF with $2TW$ degrees of freedom and a noncentrality parameter of $2E_S/N_{01}$, where E_S is the signal energy received during a time interval of length T [11]. The noncentral chi-square PDF is mathematically described as [9]

$$f(v|\text{signal} + \text{noise}) = \frac{1}{2\sigma^2} \left(\frac{v}{2E_S/N_{01}} \right)^{(TW-1)/2} e^{-(\frac{2E_S}{N_{01}} + v)/2\sigma^2} \cdot I_{TW-1} \left(\sqrt{\frac{2vE_S}{N_{01}}} / \sigma^2 \right) \quad (4)$$

where $I_x(\cdot)$ is the modified Bessel function of order x . Figure 3 illustrates an example of these conditional PDFs.

For large TW products, the output statistics approach Gaussian density functions for both the noise and signal plus noise cases. Simple detectability models make this approximation and often further assume that the variance of the noise is equal to the variance of the signal plus noise (i.e., the signal is very weak). After signal detectability is calculated, a correction factor is typically applied to correct for the error introduced by making the Gaussian assumption (see for instance, [11, pg. 298]).

LPI/SDA calculates the required C/N_{01} using chi-square statistics except in the case of large TW products (> 500) where the Gaussian approximation is very good, and makes no assumption whatsoever about the signal and noise powers. A typical detectability calculation in LPI/SDA proceeds as follows. The user describes a signal and radiometer, and specifies a TW product, P_{FA} and P_D . LPI/SDA maps P_{FA} and P_D to Q_{FA} and Q_D if necessary. The radiometer output PDF, given that the input is noise alone, is now fixed, and a detection threshold can be calculated by using a chi-square tail function routine given in [6] to yield the correct Q_{FA} . With knowledge of this threshold, the noncentrality parameter of the output PDF given signal plus noise at the input can be varied until the correct Q_D is obtained. Q_D is calculated using a recursive generalized Q function algorithm given in [2]. For certain combinations of parameters, the generalized Q function can be interpreted as the cumulative distribution function (CDF) of a noncentral chi-square random variable.

The received signal energy in time T is equal to the signal power times T

$$E_S = C \cdot T \quad (5)$$

With the radiometer output PDFs determined, the noncentrality parameter of the noncentral chi-square PDF ($2E_S/N_{01}$) can be multiplied by $1/2T$ to yield the required C/N_{01} .

LPI/SDA modifies this calculated C/N_{01} appropriately if the radiometer dimensions are not perfectly matched to the transmitted signal. This would be the case if, for example, a DS signal is specified along with a wideband radiometer whose bandwidth is something less than the spread bandwidth of the signal. For example, assume that a DS signal is incident on a wideband radiometer, and the bandwidth of the radiometer is half of the spread bandwidth of the signal. The radiometer receives only half of the energy contained in the transmitted signal; thus, the calculated C/N_{01} , which indicates the amount of signal power that the radiometer must receive, is multiplied by a factor of two (the ratio of the signal bandwidth to the radiometer bandwidth) to yield the true relative power that the signal must contain. An adjustment is also made to C/N_{01} if a time hopped signal (or any hybrid which includes time hopping) is incident on a wideband radiometer. Because the received energy is not uniformly distributed within the integration time, the required C/N_{01} must be multiplied by the inverse of the signal duty cycle $T_1/(b_2 \cdot T_2)$.

For TW products greater than 500, the calculation proceeds similarly, except that the output PDFs are assumed to be Gaussian with the first and second moments equivalent to the corresponding chi-square moments, and thus Gaussian tail probabilities are used rather than chi-square tail probabilities. The Gaussian tail probability calculations are performed using Q-function algorithms contained in [6].

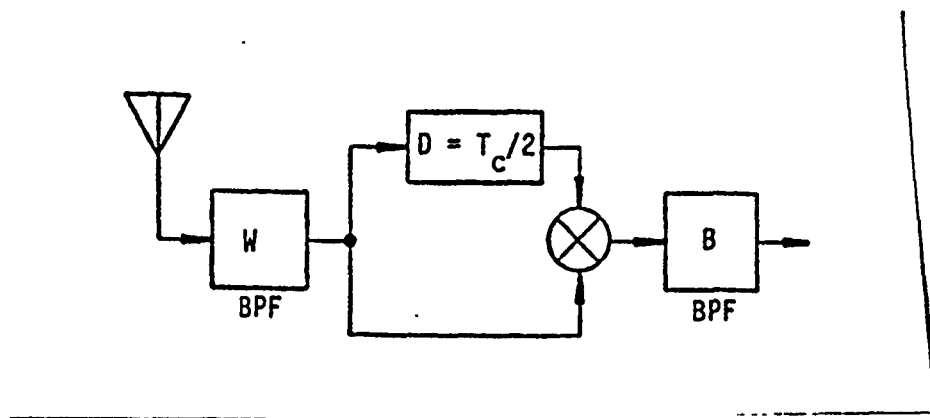


Figure 4: Delay and Multiply Chip Rate Detector

4 Feature Detectors

LPI/SDA has the ability to model three types of feature detectors. They are the chip rate detector (delay and multiply method, and envelope method) and the frequency hop rate detector. The relationship between the SNRs at the input and output of a feature detector are expressible in closed form as a function of spread spectrum modulation type and narrowband modulation type. These expressions are given in [5]. The remainder of this section briefly described the operation of the three feature detectors.

The delay and multiply type of chip rate detector, illustrated in Figure 4, filters a portion of the RF spectrum (typically the spread bandwidth of the transmitted signal), splits this filtered signal into a nondelayed path and one which is delayed by half a chip interval, multiplies these signals together, and filters the output with a narrowband filter. When the signal is present at the intercept receiver, and assuming that the chip rate is constant (not dithered), then there will appear in the spectrum of the output signal a strong spectral line at a frequency corresponding to the chip rate. The detection of this spectral line is accomplished by using a radiometer with a front end bandwidth equal to the narrowband output bandwidth of the chip rate detector and integration time equal to the transmitted message duration.

The envelope method of chip rate detection, illustrated in Figure 5, has long been used in communication systems as a means of extracting a clock signal for demodulation of PSK data. It works on the principle that a PSK or QPSK signal is a constant envelope modulation having a spectral response similar to a $\sin(x)/x$ power distribution. By centering a bandpass filter W_2 on the carrier frequency and matching the bandwidth of transmission, it then becomes possible to substantially filter the spectrum over its sidelobes. This transforms the constant envelope modulation into one which has dips or notches wherever there is a data transition. It is the detection of these dips and notches by means of an envelope detector which produces the periodic signal at the chipping rate frequency. The bandpass filter W_3 is selected to extract

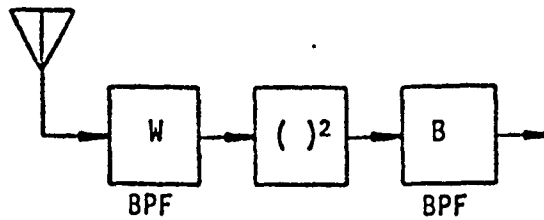


Figure 5: Envelope Chip Rate Detector

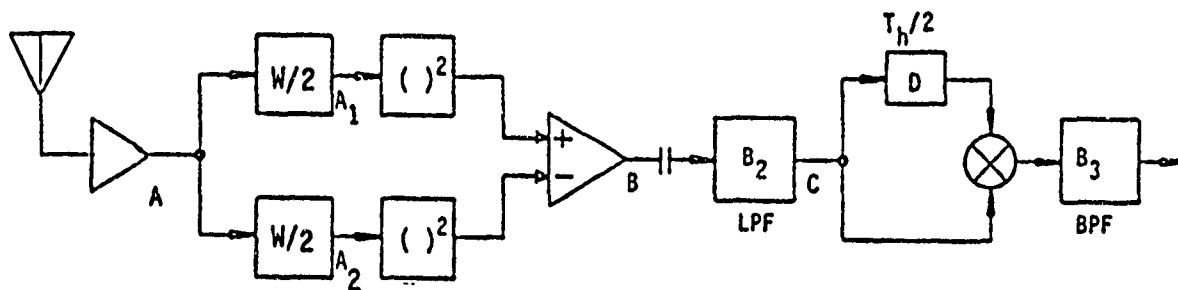


Figure 6: Frequency Hop Rate Detector

the spectral line in the output spectrum and a radiometer is used as before to detect the presence of the line.

The final type of feature detector, the hop rate detector, is illustrated in Figure 6. This system partitions the spread spectrum bandwidth into upper and lower halves, each filtered by a front-end filter. The output of each filter is passed through a square-law device and these signals are differenced. Thus, a frequency hopping signal which randomly hops into each of the front-end filters, generates a nearly square wave signal at the output of the differencer which has a strong periodic component corresponding to the frequency hop rate. This signal is then processed by a delay and multiply chip rate detector which is followed by a radiometer as before.

5 Quality Factors

Gutman and Prescott [4] describe five LPI system Quality Factors which quantify the effectiveness of an LPI communications system in the presence of jammers and intercept receivers. The material in this section explains these Quality Factors only very briefly.

A typical LPI communications scenario is depicted in Figure 7, representing any situation in which a cooperative transmitter and receiver are targeted by both jammers (which disrupt the communications receiver) and intercept receivers (which attempt to detect and exploit the transmitted signal).

The communications transmitter and receiver, jammer, and intercept receiver all have at their disposal certain techniques for maximizing their effectiveness and minimizing the performance degradation caused by other sources. The communications receiver, for example, may use null steering antennas, adaptive interference suppression filters, and employ coherent processing to counter the effects of jammers.

The relationships among the key players (communications transmitter and receiver, jammer, and intercept receiver) can be analyzed to reveal potential trade-offs that may exist by performing a simple link analysis. The link analysis reveals strengths and vulnerabilities, and provides the system designer with the insight to determine how to most effectively concentrate system resources.

This link analysis, described fully in [4], yields five Quality Factors which are typically expressed in dB. These are:

- Antenna Quality Factor $Q_{ant} = G_{tc}G_{ct}/G_{ti}G_{it}$ where G_{tc} is the transmitter antenna gain in the direction of the receiver, G_{ct} is the receiver antenna gain in the direction of the transmitter, G_{ti} is the transmitter antenna gain in the direction of the interceptor, and G_{it} is the interceptor antenna gain in the direction of the transmitter.
- Atmospheric Quality Factor $Q_{atm} = L_i/L_c$ where L_i and L_c are the losses (other than free space loss) for the interceptor and communications link respectively. These losses include gaseous attenuation, depolarization due to hydrometeors, etc.
- Adaptive Technologies Quality Factor $Q_{ada} = N_i/N_c$ where N_i and N_c are the total received noise powers at the interceptor and receiver respectively. This Quality Factor compares the ability of both the communications and intercept receivers to adaptively filter interference.
- Modulation Quality Factor Q_{mod} - this Quality Factor is the ratio of the signal to noise power spectral densities (SNRs) received at the interceptor and receiver respectively. The SNR at the communications receiver is determined by the

tolerable bit error rate, the type of narrow-band modulation used, and the margin required to overcome fading in the channel.

- LPI Quality Factor $Q_{LPI} = (R_c/R_i)^2$ where R_c and R_i are, respectively, the range from transmitter to receiver and from transmitter to interceptor. This Quality Factor is also the product of the other four Quality Factors.

6 LPI/SDA Operation and Sample Calculations

The LPI/SDA user interface contains independent input pages for describing the spread spectrum signal and radiometer. Figure 8 illustrates the structure of the user interface and the paths which one may follow when going from one input page to another. On each of these pages, the user selects a signal or radiometer type via pop-up menus, and specifies the parameters that accompany the selected type. Another page allows the user to specify P_D and P_{FA} and calculate the required C/N_{01} . On yet another page, the user may input communications link information (path losses, antenna gains, receiver noise temperatures, etc.) which are required for Quality Factor calculations. Values for the five Quality Factors can be computed, and one of these Quality Factors, the Modulation Quality Factor, can be plotted against one of six system parameters, including P_D and P_{FA} . The user may move freely from page to page to change the signal type, radiometer type, and associated parameters and find immediately the effect of these changes on the required C/N_{01} . Compatibility and range checking are performed on all input to ensure that they are tenable.

Figure 9 illustrates a plot of the calculated C/N_{01} required as a function of Q_D for several values of Q_F and a TW product of 1. Figure 10 is similar except for a TW product of 10,000. These curves correspond within .2 dB to the curves of required E_S/N_1 presented in [1, pp. 65-70] (note that $E_S/N_{01} = C/N_{01}$ when $T = 1$).

Woodring and Edell [11] present several examples of calculating the detectability of spread spectrum signals by radiometers. In their first example, they describe a frequency hopped signal with a message time $T_1 = 4$ sec, a spread bandwidth $W_1 = 2$ GHz, and a hop rate of 2000 hops/sec (which together with T_1 corresponds to $b_2 = 8000$ and $T_2 = 500$ μ sec). The desired P_D and P_{FA} are 0.1 and 10^{-6} respectively. They report a required C/N_{01} of 48.9 dB-Hz, and given the same parameters, LPI/SDA returns 48.900 dB-Hz. In their second example, they calculate the detectability of a single hop or pulse. In this case, $T_1 = T_2 = 500$ μ sec is the pulse duration and $W_1 = 2000$ Hz is the pulse bandwidth, yielding a TW product of unity. Note that $b_2 = 1$. Woodring and Edell give a required C/N_{01} of 41.7 dB-Hz whereas LPI/SDA returns 41.740 dB-Hz.

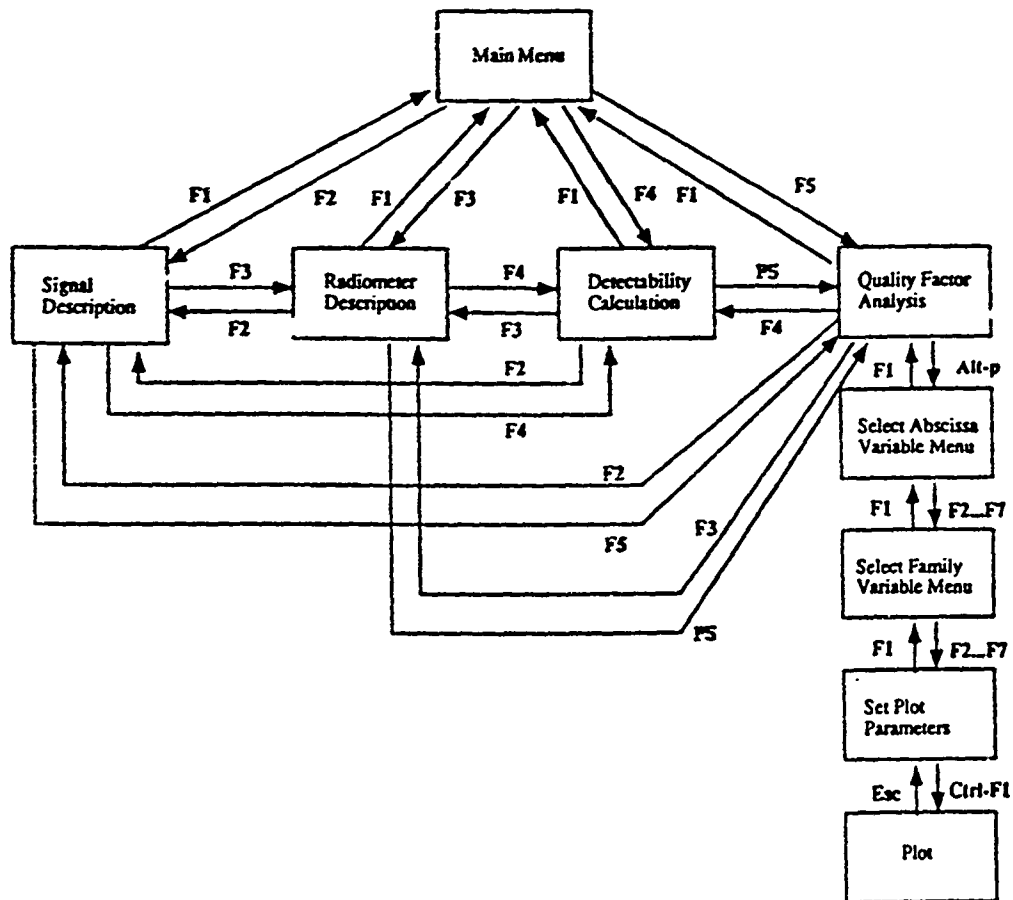


Figure 7: LPI/SDA user interface structure.

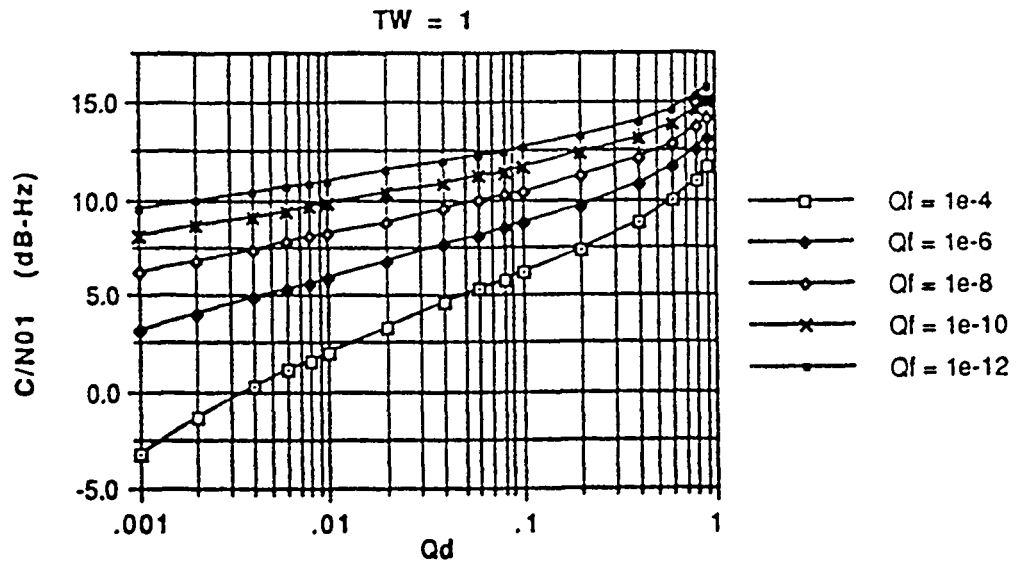


Figure 8: Required C/N_{01} for $TW = 1$.

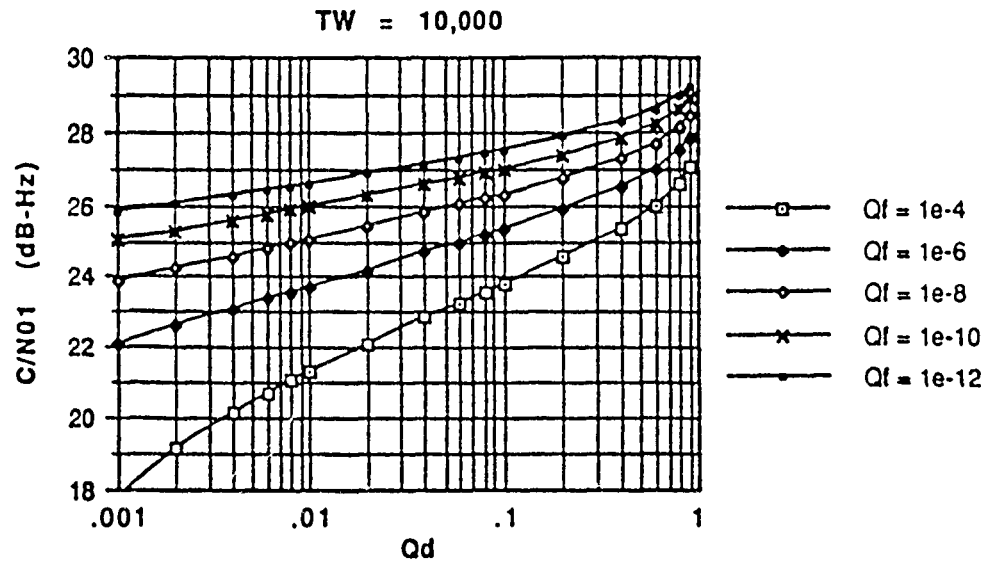


Figure 9: Required C/N_{01} for $TW = 10,000$.

7 Conclusions

LPI/SDA is an efficient and highly accurate computer aided analysis tool to assist the LPI communications design engineer in determining the detectability of certain spread spectrum waveforms by both radiometric and feature detectors. Using this tool, the designer can construct a signal waveform which is least vulnerable to detection by the most likely detection threat. Further, the designer can analyze the covertness of the LPI signal in terms of scenario dependent and scenario independent factors via calculation of five different Quality Factors.

LPI/SDA's results have been shown to be within .2 dB of similar analytical results presented in [1] and [11].

References

- [1] Dillard, George M. and Robin A., "Detectability of Spread-Spectrum Signals," Artech House, Norwood, MA, 1989.
- [2] Dillard, George M., "Recursive Computation of the Generalized Q Function," IEEE Transactions on Aerospace and Electronic Systems, July 1973, pp. 614-615.
- [3] Dillard, Robin A., "Detectability of Spread- Spectrum Signals," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-15, No. 4, July 1979.
- [4] Gutman, Lawrence L. and Prescott, Glenn E., "System Quality Factors for LPI Communications," Proceedings of the 1989 IEEE International Conference on Systems Engineering, Dayton, Ohio, 1989.
- [5] Nicholson, David L., "Spread Spectrum Signal Design: LPE and AJ Systems," Computer Science Press, Rockville, MA, 1988.
- [6] Press, William H., Flannery, Brian P., Teukolsky, Saul A., Vetterling, William T., "Numerical Recipes: the Art of Scientific Computing," Cambridge University Press, New York, 1986.
- [7] Simon, M., Omura, J., Scholtz, R., and Levitt, B., "Spread Spectrum Communications," Vol. 3, Computer Science Press, Rockville, MA, 1985.
- [8] Skolnik, Merrill I., "Introduction to Radar Systems," McGraw-Hill, New York, 1980.
- [9] Torrieri, Don J., "Principles of Secure Communication Systems," Artech House, Norwood, MA, 1985.
- [10] Urkowitz, Harry, "Energy Detection of Unknown Deterministic Signals," Proceedings of the IEEE, Vol. 55, No. 4, April 1967.

- [11] Woodring, D. and Edell, J., "Detectability Calculation Techniques," U.S. Naval Research Laboratory, Code 5480, Sept. 1, 1977.

MINIMUM RESOLVABLE TEMPERATURE MEASUREMENTS OF THE MITSUBISHI THERMAL IMAGER IR 5210C

Darron D. Lockett, Research Associate
Electro-Optical Sensor Evaluation/Analysis Group
Wright Patterson Air Force Base, Ohio

ABSTRACT

The Electro-Optical Sensor Evaluation/Analysis Group of Wright Patterson Air Force Base has been involved in extensive modeling of mid-wave infrared sensor systems. One study that is of particular interest is the characterization of sensor performance. The Minimum Resolvable Temperature (MRT) test is used as a figure of merit upon which to judge imaging infrared performance. In this report MRT's of the prototype Mitsubishi Thermal Imager IR 5210C are measured. The result of this test is a plot of the minimum target to background temperature differences that allow humans to resolve a series of varying spatial frequency targets. These measurements may be used as inputs for the Mid-Wave Infrared Sensor Model developed by the Group.

1.0 INTRODUCTION

There has been a recent revival of interest within the Air Force in imaging mid-wavelength infrared (MWIR) sensor technology. This increased interest in MWIR has resulted from advances made in detector array technology and fabrication. Although non-imaging air-to-air infrared systems have operated in the 3-to 5-micron waveband for some time, most imaging air-to-ground infrared sensors have operated in the 8-to 14-micron

(long-wavelength infrared) waveband. However, prototype systems have been designed and built using MWIR sensor technology. These systems are now being proposed for air-to-ground target detection and recognition applications.

The Electro-Optics Branch of the Air Force's Avionics Laboratory has been involved in detailed sensor performance modeling and analysis activities for current and future electro-optical sensor systems. In response to the increased interest in MWIR sensor technology, the Electro-Optics Branch and Battelle's Columbus Division have developed the Mid-Wave Infrared (MWIR) Sensor Model, which provides the Avionics Laboratory with extensive modeling capability.

The MWIR Sensor Model is a computer model intended for use in studying mid-wave (3-to 5-micron) staring focal plane arrays. This includes HgCdTe, PtSi, and generic background limited infrared detectors. The model was designed for use in all components of detector performance, including noise equivalent temperature differences and modulation transfer functions (MTF). The model also calculates the ranges at which detection and recognition of complicated and generic targets can be accomplished. This information is used to assess the quality of the sensor and its effectiveness in any given operational scenario. The Air Force is interested in comparing all kinds of sensor systems in all kinds of mission scenarios so that the right electro-optical systems are used for the right situation.

Each component level of the MWIR Model has tremendous breadth. Therefore, a detailed discussion of the model is beyond

the scope of this report. However, the concept of the Minimum Resolvable Temperature (MRT) is explored because of the information it provides on both thermal sensitivity and spatial resolution of the sensor. Actual Laboratory tests for MRT's of the prototype Mitsubishi Thermal Imager IR5210C are performed and may be used as inputs to the MWIR Sensor Model.

The Mitsubishi Thermal Imager IR-5120C is an instrument that captures the 3- to 5-micron infrared rays emitted from an object and converts the rays into a real time image. It uses a 512 X 512 Silicide Schottky Barrier focal plane array, and can produce a clear image of 1/60 sec field time. The NETD (Noise Equivalent Temperature Difference) is 0.15° celcius. The field of view is 14° X 11°.

The Mitsubishi Thermal Imager consists of a camera head and a camera controller. The main components of the camera head are the lens system, an infrared sensor, and a coder. The infrared rays from the object pass through the air and the lens system, and are projected on the surface of the detector. The detector consists of two dimensional arrayed elements. Each element generates an electric signal by opto-electronic conversion. The level of charged signal corresponds to the amount of infrared rays received by the element. The charged signals are sequentially read out from each element. The camera controller receives the detector output from the camera head, amplifies the signal, compensates the fixed noise pattern due to the characteristics of element, and produces an image signal.

2.0 LABORATORY MRT TESTING

The standard MRT test uses a four bar target with equal bar and space widths. The height of the target is equal to the width of the four bars and three spaces combined. This gives each bar an aspect ratio of seven to one and a total target aspect of one to one. The frequency (cy/mr) of the target is the reciprocal of the angle (mr) subtended by the projected image of a bar/space pair (1 cycle). Targets of different spatial frequencies are obtained by varying the width of the bar and space combination. The target can be oriented with the bars vertical or horizontal. By using vertical bars, the resolution in the horizontal direction can be tested. Therefore, vertical bar targets are used to produce horizontal MRT curves and horizontal bars are used to produce vertical MRT curves.

The target source consists of a target plate and a blackbody source. The target plate is placed in front of the blackbody source. It is typically constructed of metal and finished so that it has a very high emissivity. The targets are made by cutting bars out of the plate. When the target source is imaged by the imager, the plate temperature appears as the target. There are generally multiple targets of different spatial frequencies on a single target plate. The targets are arranged on the plate so that one frequency at a time can be viewed. The temperature of the blackbody source can be controlled allowing the temperature difference between the plate and source to be varied. The plate is typically kept at ambient temperature, although its temperature can also be varied. The target source

produces a thermal image of a four bar target in which the frequency and the temperature difference can be varied.

The target is typically imaged using a collimator and presented to the Imager. The collimator focuses the image to simulate a target at infinity. This allows the Imager optics to be focused at infinity. However, a collimator was not used in this report due to the inability to focus the target.

The target is imaged by the Imager and observed on a video monitor. In the standard laboratory test, only one target is presented to an observer at a time. The observer starts with a low frequency target at a temperature difference of zero degrees. He increases the bar temperature in very small increments until the target is just resolvable, bright bars on a dark background, and records this temperature (positive MRT). He then decreases the bar temperature until the target is again just resolvable, dark bars on a bright background, and records this temperature (negative MRT). The average of the absolute temperature differences is the minimum resolvable temperature difference for that spatial frequency. The target frequency is changed and the test repeated. The frequency is increased incrementally until the observer can no longer resolve the target.

The MRT data points are plotted on a graph of frequency versus minimum resolvable temperature difference. This plot is the MRT curve for the tested sensor. The MRT curve will go asymptotic at the system cutoff frequency. The sensor will have two MRT curves, one horizontal and one vertical.



In this report the sensor is situated 15.75 inches from the target plate. The electronic control panel consists of a temperature delta selection switch, which allows temperature increments as small as 0.01 degree celcius, and a hot/cold target selection switch, which reverses the direction the black body source drives to the selected temperature delta. The electronics sense the temperatures of both the target plate and blackbody source, and drive the blackbody source temperature to obtain the selected target to background temperature delta.

Since the test involves human judgement to determine resolution, there can be many different curves for a single target. By using multiple observers and statistical averaging of the data sets, the repeatability of the results can be increased.

3.0 PLOTS OF MRT VS. SPATIAL FREQUENCY

The following are plots of data taken by different observers under the same laboratory conditions. Figures 1 and 2 are plots of the individual observer's data. Figure 3 is a plot of all observers' data. Figure 4 is a plot of the average horizontal and vertical MRT measurements.

Mitsubishi MRTs

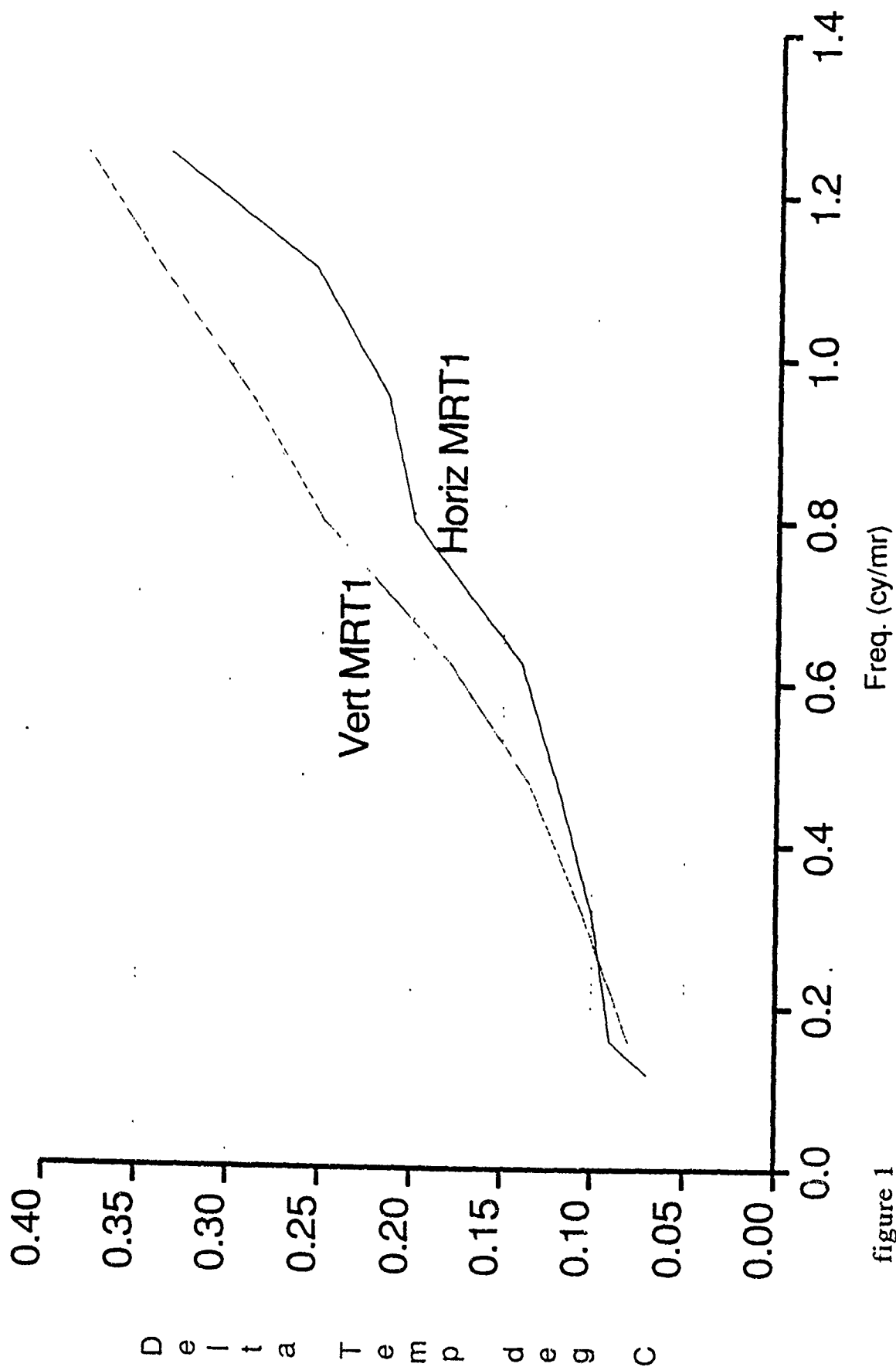


figure 1

Mitsubishi MRTs

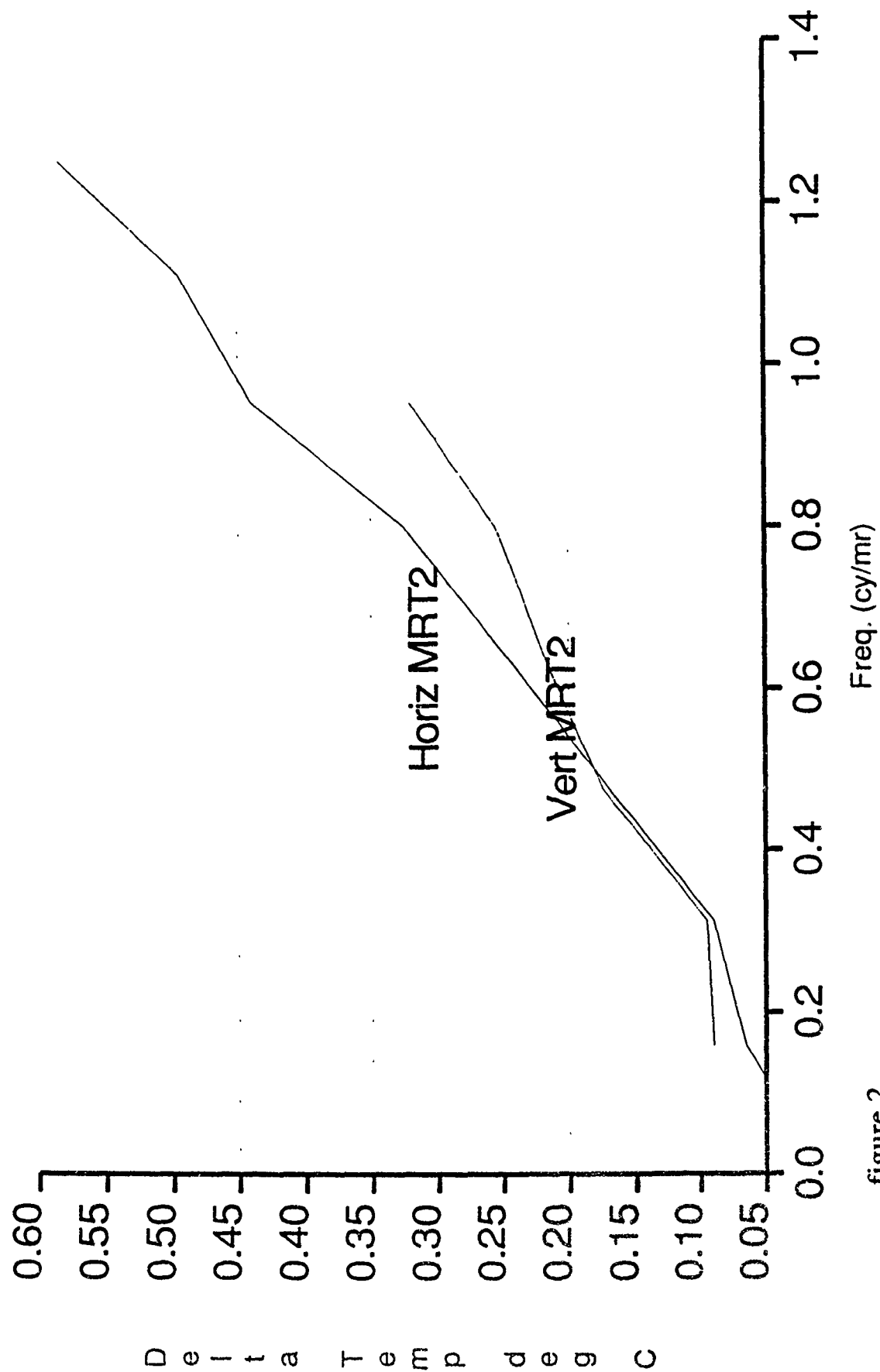


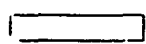
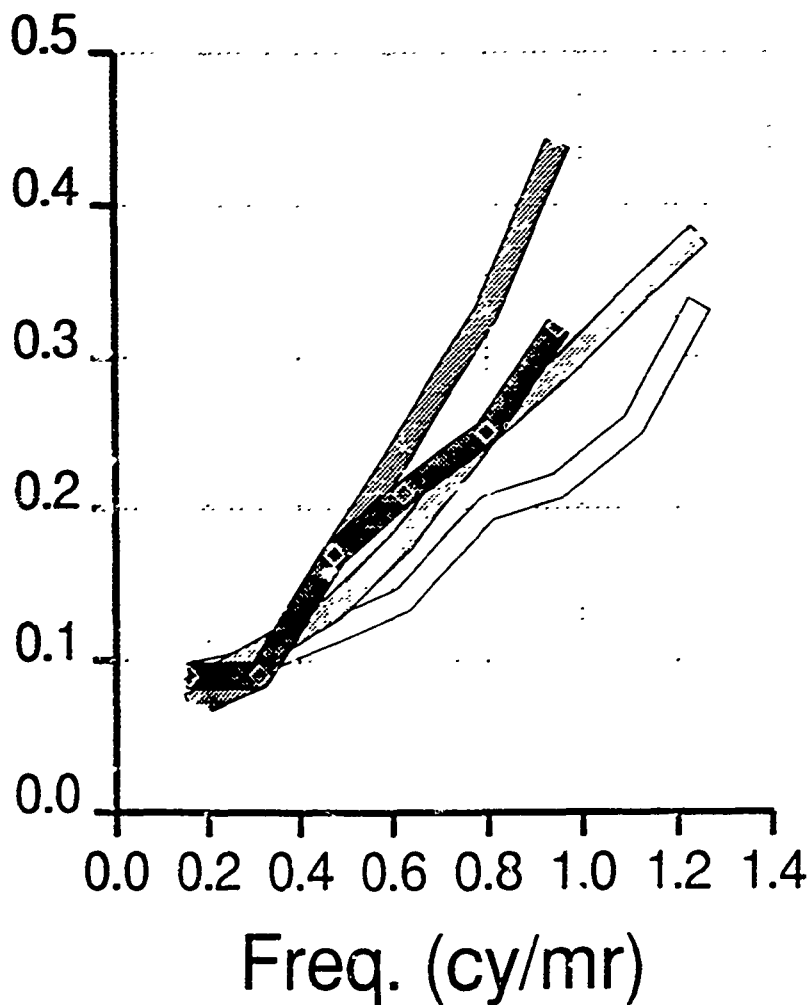
figure 2

D
e
l
t
a

T
e
m
p
e
r
a
t
u
r
e

figure 3

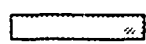
Mitsubishi MRTs



HORIZ 1



HORIZ 2



VERT 1



VERT 2

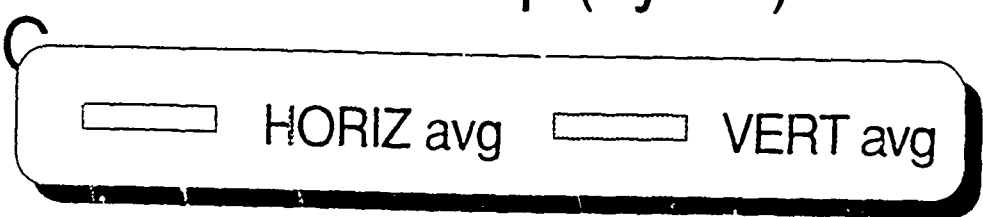
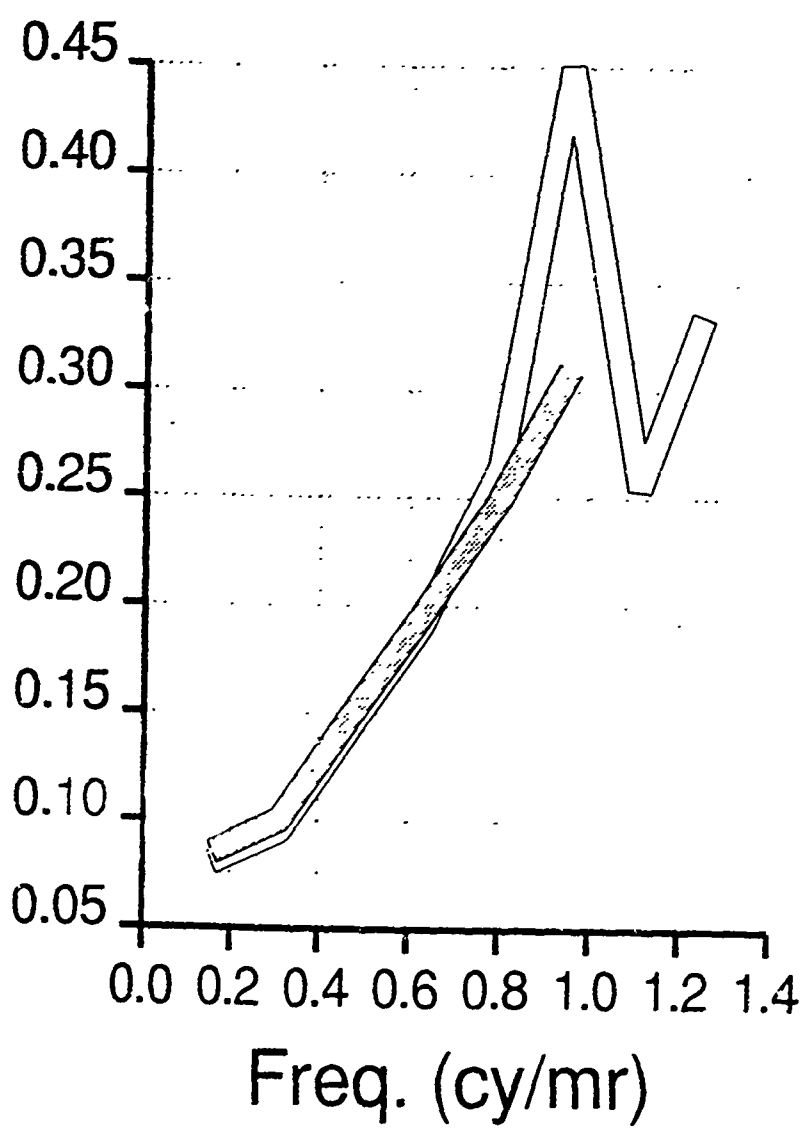
D
e
l
t
a

T
e
m
p

d
e
g

figure 4

Mitsubishi MRT's





[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

4.0 CONCLUSION

The horizontal and vertical MRT's correlate well for the lower spatial frequencies. The deviation occurs at about 0.85 cy/mr. Here the horizontal response diverges, and the vertical response data points cease to exist. The horizontal bars are not resolvable for frequencies greater than 0.9 cy/mr.

There are several concerns about the accuracy of this test. Since this test was conducted without the use of a collimator, it is questionable whether or not the measurements are valid estimates of MRT's under optimum conditions. After consulting colleagues, one opinion is that the test is legitimate because of the fact that the imager senses the same thermal energy of the target without the collimator as it does with the collimator. A contrary opinion is that the target is refocused with the use of a collimator and the resolution is altered, thereby sometimes requiring a new delta temperature to successfully resolve the target.

There is also a discrepancy in the four bar pattern target. When the smallest targets are magnified and viewed under a scope, imperfections are noticeable. The edges of the bars appear to be rough, thereby affecting the observers judgement. This may account for the divergence and inexistence of data points of the plots. A photograph of the imperfect target is given.

Future goals will be to test the validity of the procedure presented here by comparing this data with that taken during a test which uses a collimator. However, further study is hindered until the target plate is corrected.

REFERENCES

- 1) Richard Hudson Jr., Infrared Sensing Engineering, 1969.
- 2) "Mid-Wave Infrared Sensor Performance Model Analyst's Manual", Battelle Columbus Division, 1989.
- 3) Bob Muse, Rich Hill, and Fred Blommel; Electro-Optical Sensor Evaluation/Analysis Group; Personal Conversations.

Performance Evaluation of Rule Grouping Algorithm Running on the Activation Framework Architecture

Ing-Ray Chen and Bryant Poole*

August 9, 1991

Abstract

The objective of this research is to design and develop rule grouping algorithms for maximizing the performance of real-time rule-based systems running on the Activation Framework (AF) architecture [GRE 87]. This research involved the development of a formal theory for quantifying the estimation of performance metrics based on probabilistic parameter modeling and the design of a rule grouping algorithm based on Kernighan-Lin (KL) heuristic graph-partitioning for a single processor system. A demonstration system based on the theory and algorithm has been developed and tested on a portion of Advanced GPS Receiver (AGR) and Manned Maneuvering Unit (MMU) knowledge bases.

1 Introduction

The SKRAM toolset is an ongoing project [GRE 91A] that is designed to automate the development and testing of real-time expert systems running on the Activation Framework (AF) architecture. The AF architecture supports a dynamically prioritized, distributed, message based, object oriented, real-time Ada run-time environment. The user is allowed to specify independently scheduled groups of rules which are automatically converted into Ada Activation Framework Objects (AFOs). Unlike a conventional real-time scheduler which uses a static time-sliced scheduling mechanism to execute AFOs in a processor, the AF architecture uses an innovative event-sliced scheduling mechanism: the AFO priorities are changed dynamically by the arrival of events (e.g., messages sent from one AFO to other AFOs or alarm messages sent from the AF to AFOs) which reflect the change of the system environment; the most important AFO is always scheduled to execute next.

*The authors are with the Department of Computer and Information Science, University of Mississippi, Weir 302, University, MS 38677

The advantage of this approach is that it eliminates the use of interrupts to invoke the processing of critical AFOs. In the AF architecture, the importance of AFOs is always re-computed every time an event occurs, e.g., when a message carrying a new fact is delivered by the AF. Furthermore, this guarantees that the system runs in real-time because it ensures that the most important rule or AFO is executing at any time.

Despite the technical soundness of the SKRAM project for real-time expert systems, there is a perceivable system overhead for executing the event-drive scheduling mechanism. First, the processor is essentially interrupted (to determine the most important AFO) every time a message is delivered from one AFO to other AFOs, or when an alarm is triggered. Second, if the currently executing AFO indeed becomes less important than some other AFO, then the AF must perform a context switch to allow the most important AFO to run. This issue is further complicated by the rule-grouping algorithm which allocates independently scheduled rules to separate AFOs. For example, if each rule is allocated to a separate AFO, then this overhead may be quite significant because the AF has to deal with a handful of processes (AFOs) at the AFO level and the chance that this overhead will become a factor is high; on the other hand, if the rule-grouping algorithm allocates many rules to a single AFO, then, although the overhead at the AFO level is reduced (because there are fewer processes), the overhead required for scheduling different rules within a single AFO may again be significant. A similar mechanism (e.g., [GRE 91B, FOR 82]) has to be applied at the rule level for scheduling the execution of rules within a single AFO, thus introducing a lot of system maintenance overhead for keeping track of which rules are ready and which rule should be fired, etc. Both overheads, one at the AFO level and the other at the rule level, degrade the performance of the system since they defer the time required for firing productions rules. Therefore, the best rule-clustering algorithm for optimizing the system performance should balance these two overheads, thereby minimizing the total system overhead.

Current research directions toward rule-clustering algorithms for expert systems [MEH 91, JAC 90] are not tied in with the AF architecture and thus do not consider this design trade-off. The basic approach of these algorithms is to select a distance metric between each pair of rules and, after allocating each rule to a separate group, iteratively merge groups with the minimum inter-group distance until a stopping condition is met. These algorithms stop either after a pre-determined number of groups is obtained [MEH 91] or the inter-group distance between the next two groups to be combined is no longer positive [JAC 90]. These algorithms are not applicable to real-time expert

Manned Maneuvering Unit (MMU) knowledge bases and compare the theoretical prediction with the empirical result for the AGR system. Finally, Section 6 summarizes the report and outlines some future research areas.

2 Modeling of AF Architecture

In the construction of the model for describing the run-time behavior of the AF architecture, we first distinguish the following two classes of AFOs: (a) Application Code Objects (ACOs) which are created from groups of production rules, and (b) System Code Objects (SCO) which perform system-level functions such as injecting or logging data [RTI 91].

A Markov probability model is proposed as a tool for evaluating the performance of rule clustering algorithms running on the single-processor AF architecture. The model uses a finite state machine in which the system execution is modeled as a progression through the following states:

- *ACO*: in this state the CPU is allocated to an ACO. The ACO may be performing useful work, e.g., firing a rule, generating some new fact, etc., or it may be spent for the rule-level overhead, e.g., determining which rules are ready and which rules should be fired by applying a pattern matching algorithm for saving the states of the rules within a single ACO, etc.
- *AF*: in this state, the CPU is allocated to the AF's operating system. This state models part of the AFO-level overhead due to the employment of the event-driven scheduling mechanism. If the AF gets control because the currently executing ACO sent a message to a value distribution node (VDN) [RTI 91], then the CPU power is used to (a) deliver the message, (b) calculate the priming function of the receiving ACOs, and (c) compare the importance levels of the sending and receiving ACOs to determine which ACO is to get the CPU next.
- *SW*: in this state the AF uses the CPU for performing a context switch, i.e., saving the states of the sending ACO (or the ACO which just returned to AF) and transferring the control to the ACO having the highest importance level. This state also models part of the AFO-level overhead.
- *SCO*: in this state an input or output SCO takes over due to a data injection or logging event (e.g., an alarm event). The CPU power is used to service the I/O alarm event, switch the CPU from the currently executing ACO, perform a data injection or logging operation, and

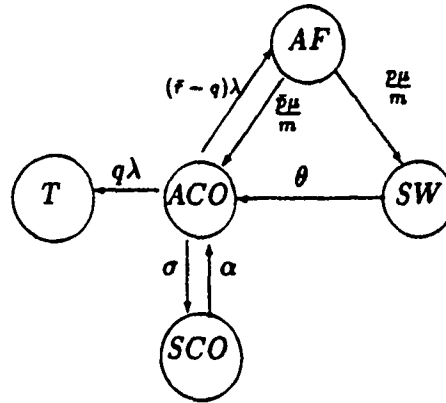


Figure 1: The Markov Model for the Single Processor AF Architecture.

switch the CPU back to an ACO when the operation is completed. This state also models part of the AFO-level overhead.

- T : the termination state.

Figure 1 shows the model which is constructed as follows:

1. When the system is in the ACO state, the CPU is used by an ACO for processing an arriving fact, i.e., determining which rules are ready to fire due to the arrival of the new fact, and selecting one to fire among the set of ready rules, if any. Firing a rule may generate new facts which may (a) terminate the system, (b) instantiate other rules within the ACO, thus requiring the ACO to process the new fact again, or (c) instantiate rules in other ACOs, thus requiring the ACO to send a message carrying the fact to other ACOs via a VDN. The time that is required for the ACO to process an arriving fact is assumed to be exponentially distributed with a constant rate λ . The probability of terminating the system (event (a) above), executing the same ACO again (event (b)) and returning the control to AF (event (c) or when there is no rule that can be fired) are q , r , and $1 - r - q$, respectively. In Figure 1 these correspond to the horizontal transition at rate $q\lambda$, the self-looping transition at rate $r\lambda$ (not shown in the figure) and the diagonal transition at rate $(\bar{r} - q)\lambda$, respectively.
2. When the AF takes over (i.e., in state AF), the time that is required for the AF to deliver a message from the sending ACO to another ACO and re-compute the priorities of these two

ACOs to determine which one has the higher priority is assumed to be exponentially and randomly distributed with a constant service rate μ . Assume that the average number of ACO connections for each rule having at least one ACO connection is m . Then, the overall rate for the AF to deliver a message and re-compute priorities of all involved ACOs is $\frac{\mu}{m}$. Further, it is assumed that with probability p , the control will be transferred to a new ACO. Hence, with probability \bar{p} the sending ACO will retain the CPU. These events correspond to the diagonal transitions at rate $\frac{\bar{p}\mu}{m}$ if the CPU is retained by the sending ACO and at rate $\frac{p\mu}{m}$ if the CPU is allocated to a new ACO, respectively.

3. When the AF allocates the CPU to a new ACO, the system is in the *SW* state. The CPU time is used for performing a context switch. The time for performing a context switch is assumed to be exponentially and randomly distributed with a constant context-switching rate θ . After the context switch is performed, the CPU is allocated to a new ACO and the system is again in state *ACO*. This event corresponds to the horizontal transition at rate θ .
4. If the list of SCOs to be checked by AF at regular intervals exists, then state *SCO* exists. (See section 4, AFC manual [RTI 91].) In this case, whenever an ACO makes a call to a framework function, AF will compare the importance level of the executing ACO with those SCOs in the list (if sufficient time has elapsed since the last time the comparison was performed) to see if the CPU should be allocated to one of them. The time for this event to occur is assumed to be exponentially distributed with a constant alarm rate σ (e.g., the data sensor rate plus data logging rate). The event corresponds to the vertical transition at rate σ from state *ACO* to state *SCO*. When the system is in state *SCO*, the time required for the system to return to state *ACO* is assumed to be exponentially distributed with a constant rate α . This quantity includes the time required to perform a data injection or logging operation as well as the time for performing a context switch twice: one to switch from state *ACO* to *SCO* and one to switch back from *SCO* to *ACO*. This event corresponds the vertical transition at rate α from state *SCO* to state *ACO*.
5. The system continues to perform state transitions until it reaches state *T*.

The performance measure of interest in this model (Figure 1) is the mean time to termination (*MTTT*), i.e., the mean time required for the system to transit to state *T* from state *ACO*. This

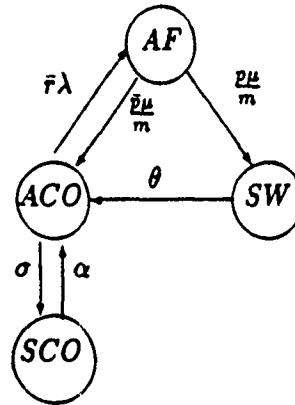


Figure 2: The Single Processor Markov Model with No Termination State.

measure reflects the average time required for a rule system to terminate. Thus, the design goal in this case is to minimize the *MTTT*.

A variation of the above model is the case in which a rule system never terminates (e.g., running an infinite loop in an embedded process-control system). This is shown in Figure 2. In this case, the performance measure of interest is the effective production rate, namely, $\lambda P_{ACO}(\infty)$. This performance metric measures the number of facts generated (or rules fired) per time unit and is obtained as follows: λ is the production rate when the system is at state *ACO*, but the system only stays in state *ACO* with probability $P_{ACO}(\infty)$; consequently, the effective production rate is $\lambda P_{ACO}(\infty)$. Note that this performance metric accounts for the trade-off between the AFO-level and rule-level overheads by considering the proportion of time the system stays at state *ACO* (which accounts for the AFO-level overhead) and the rate at which the system fires a rule once the system is in that state (which accounts for the rule-level overhead). An important observation is that λ decreases as more rules are grouped within one ACO because more time is spent for processing a new arriving fact over a larger sorting network.

The following table summarizes the symbols which will be used in the report.

N : number of rules in the system.

k : number of ACOs (groups) in the system.

m : the average number of external ACOs that a rule connects to.

s_j : number of rules in the j th group.

λ : the *new-fact generation* (or rule-firing) rate when the system is at state *ACO*.

λ_i : λ with each *ACO* containing i rules.

μ : the *service rate* of AF.

θ : the *context switching* rate of AF.

σ : the *alarm* rate of SCOs.

α : the service rate of an SCO, including the context switch overhead.

D : the average time to complete an I/O operation.

q : the probability that when an ACO generates a fact, the termination condition has been met.

\bar{q} : $1 - q$

r : the probability that when an ACO generates a fact, the fact will instantiate rules within the same ACO.

\bar{r} : $1 - r$

$\bar{r} - q$: the probability that when an ACO generates a fact, the fact will instantiate rules in other ACOs.

η : the average number of rules affected (that have to be inspected) in an ACO when a new fact arrives.

η_j : η with each ACO containing j rules.

p : the probability that AF transfers the CPU to a new ACO after servicing a message delivery.

\bar{p} : $1 - p$, the probability that the sending ACO retains the CPU.

$P_i(t)$: the probability that the system is at state i at time t .

p_o : the probability that the AF will allocate the CPU to a connecting ACO.

$P_i(\infty)$: the steady state probability that the system is at state i .

P_{system} : the effective production rate of an expert system, i.e., $\lambda P_{ACO}(\infty)$.

$P_{system,i}$: P_{system} with each ACO containing i rules.

Without loss of generality, the analysis henceforth assumes that the rule system never terminates. The performance measure of interest is the effective production rate, i.e., the number of facts generated (or rules fired) per second, $\lambda P_{ACO}(\infty)$, or, P_{system} . This measure can be obtained by solving for $P_{ACO}(\infty)$ from the following set of linear equations describing the Markov model in Figure 2 [ROS 89]:

$$P_{ACO}(\infty) + P_{SCO}(\infty) + P_{SW}(\infty) + P_{AF}(\infty) = 1;$$

$$\mu P_{AF}(\infty) = \bar{r} \lambda P_{ACO}(\infty);$$

$$\theta P_{SW}(\infty) = p \mu P_{AF}(\infty);$$

$$\alpha P_{SCO}(\infty) = \sigma P_{ACO}(\infty).$$

This yields

$$P_{ACO}(\infty) = \frac{1}{1 + \frac{\sigma}{\alpha} + \frac{\lambda \bar{r}}{\mu} + \frac{\lambda \bar{r} p}{\theta}}$$

and thus

$$P_{system} = \frac{\lambda}{1 + \frac{\sigma}{\alpha} + \frac{\lambda \bar{r}}{\mu} + \frac{\lambda \bar{r} p}{\theta}} \quad (1)$$

This performance equation has an important implication: for a given k -way partition (e.g., $k = 1$ means there is only one group), the system performance increases as λ , θ , μ or α increases, and as \bar{r} , p , m , or σ decreases. To improve the performance of the system, we like to decrease \bar{r} , p and m by selecting a lower k value (thus minimizing the AFO-level overhead); however, this would decrease λ (i.e., increasing the rule-level overhead) which will adversely degrade the system performance. Therefore, the goal is to select a best k value which can balance these two opposite effects.

3 Parameterization of Model Parameters

Equation (1) can serve as a basis for predicting the performance of the system when all parameters are quantified after a particular rule grouping algorithm has been applied to partition the production rules into groups. Some of these parameters are machine-dependent but insensitive to the use of rule grouping algorithms. These are called *statistically measurable* parameters. Others are sensitive to the utilization of rule grouping algorithms and the way a group processes a new fact. These are called *computable* parameters. This section discusses techniques for quantifying these two types of parameters.

3.1 Statistically Measurable Parameters

Statistically Measurable Parameters include θ , μ , σ and α . These are discussed in the following.

- θ denotes the average number of times the CPU is capable of performing a context switch (dedicated for that purpose) per time unit. For an 80386 machine, for example, the average time required for the AF to perform a context switch is about 2 milliseconds [RTI 91].

Therefore,

$$\theta \approx 500 \text{ sec}^{-1} \quad \text{for 80386}$$

- μ denotes the average number of times the CPU is capable of (a) delivering a message from an ACO to another ACO and (b) re-computing the priorities of these two ACOs to determine which one has the higher priority. For an 80386 machine, the average time required for the AF to perform this service is about 1.5 milliseconds [RTI 91]. Therefore, μ can be estimated by

$$\mu \approx 700 \text{ sec}^{-1} \quad \text{for 80386}$$

- σ denotes the average number of times per time unit that the execution of the rule system is interrupted by data injection or logging activities. For example, if it is measured that there are about five data injection and five data logging operations per second, then $\sigma = 10 \text{ sec}^{-1}$.
- α denotes the average number of times the CPU is capable of performing a data injection or logging operation, given that an alarm event has just interrupted the execution of the rule system. The way the system is modeled assumes that whenever the system is interrupted by an alarm event, an input or output SCO is ready to perform an I/O operation. Further, the time required to service the operation includes the overhead for the system to switch back and forth between states ACO and SCO. Based on this, α can be approximated by

$$\alpha \approx \frac{1}{\frac{2}{\mu} + \frac{2}{\theta} + D}$$

where D stands for the average time actually required to input/output the data from/to the I/O devices, e.g., reading a block of data from the serial port or dumping a block of data to the disk, etc. D is a statistically measurable quantity, e.g., $D \approx 20$ milliseconds for an 80386 disk read/write operation.

3.2 Computable Parameters

Computable parameters include r , m , p and λ . These parameters are sensitive to the use of rule grouping algorithms. (In fact, they are outputs of rule grouping algorithms.) In addition, λ is sensitive to the mechanism with which each group (or ACO) processes new arriving facts.

We first define some data structures used by rule grouping algorithms. Then, we explain how to compute these parameters using these data structures. Let $A[1..N, 1..N]$ be the adjacency matrix

of the input Evidence Flow Graph (EFG) that shows the connectivity of production rules, i.e., $a_{ii} = 0$ for all i and $a_{ij} = 1$ if rule j uses a fact generated by rule i ; 0 otherwise. Let $B[1..N]$ be the output partition vector of a particular rule grouping algorithm which has been applied to partition the rule production rules into k groups, such that if $b[i] = j$ then rule i is allocated to the j th group. In the following we discuss techniques for computing r , p , n and λ .

- r denotes the probability that when an ACO generates a fact, the fact will instantiate rules within the same ACO, rather than instantiating rules in other ACOs. We can approximate r by

$$r = \frac{C_i}{C} = \frac{\sum_{i=1}^N \sum_{\substack{j=1 \\ b[j]=b[i]}}^N a_{ij}}{\sum_{i=1}^N \sum_{j=1}^N a_{ij}} \quad (2)$$

where C stands for the total number of connections in the connection matrix and C_i stands for the total number of internal connections for all k groups after the partition. When there are fewer groups, $r \rightarrow 1$ and, conversely, when there are many groups, $r \rightarrow 0$. For a given k , a rule grouping algorithm should maximize r (thereby minimizing \bar{r}) as much as possible to reduce the AFO-level overhead.

- m denotes the average number of external groups that a rule connects to. This can be estimated in three steps: (a) computing, for each rule having at least one external connection, the number of groups it connects to, (b) accumulating the total count for all such rules, and (c) dividing the total count by all such rules. Hence, the average number of external ACO connections per rule is an output of a partition. Obviously, m increases as the number of groups increases and, as a result, the AFO-level overhead involved for sending a message from an ACO to all its connecting ACOs increases, a behavior which we have modeled in the Markov model.
- p represents the probability that the AF will perform a context switch after it had serviced a message delivery. If we assume that the AF, when servicing a message delivery from an ACO, inspects each connecting ACO in succession to determine whether the CPU should be switched to that connecting ACO with a success probability p_o , then the probability that the

AF will perform a context switch is proportional to m as follows:

$$p = \sum_{i=1}^m p_o (1 - p_o)^{i-1} \quad (3)$$

This approach gives a reasonable estimation of p since it can well explain why $p \approx 1$ when there are many groups.

- λ denotes the rate at which a new fact is generated by a group. The magnitude of this parameter largely depends on how many rules there are in each group. For that purpose, let λ_j denote this rate when there are j rules in each group. Furthermore, let λ_j relate to λ_1 (the new fact generation rate when each ACO contains exactly one rule) by the following equation:

$$\frac{\lambda_1}{\lambda_j} = \eta_j$$

where η_j stands for the number of rules that have to be processed by a group containing j rules when a new fact arrives. The advantage of relating λ_j with λ_1 is that λ_1 is a measurable quantity and, therefore, if η_j can be estimated properly, λ_j can be computed from λ_1 . For example, for 80386 machines with 1Mflops processing capability, if each rule contains approximately 500 machine language (floating point) instructions, then

$$\lambda_1 \approx 2000 \text{ sec}^{-1}$$

To properly estimate η_j , we distinguish the following two mechanisms with which each group may process an arriving fact:

- *Non Data-Driven Mechanism*: In this scheme, all j rules within a group are encoded as a succession of if-then code blocks. Therefore the processing time required to generate a new fact is about j times as large as required when there is only one rule per group because all j code blocks must be examined before one is selected to fire. Under this *non data-driven* scheme:

$$\eta_j = j ; \text{ and } \lambda_j = \frac{\lambda_1}{j}$$

Consequently, for a selected k -way partition, since $j = N/k$,

$$\lambda = \frac{\lambda_1 k}{N} \quad \text{for non data-driven}$$

where N is the number of rules in the rule system.

- **Data-Driven Mechanism:** In this scheme, each group maintains a sorting network (consisting of RTGO, BFO and RO data structure objects - see [GRE 91B]) with a size proportional to the number of rules in each group. When a new fact arrives, instead of inspecting all j rules, only a portion of these rules in the sorting network (that are affected by the new fact) are inspected before the most important rule is selected to run. As a result, η_j can be estimated by the average number of internal connections per rule, i.e.,

$$\eta_j = \frac{C_i}{N}$$

where C_i stands for the total number of internal connections for all groups as defined in equation (2). Recall that $r = C_i/C$ where C stands for the total number of connections of the input connectivity matrix. Therefore the average number of rules that have to be inspected, η_j , is related to r by

$$\eta_j = \frac{rC}{N}$$

Consequently, for a selected k -way partition,

$$\lambda = \frac{\lambda_1 N}{rC} \quad \text{for data-driven}$$

4 The Best- k Rule Grouping Algorithm

In this section, we develop a rule grouping algorithm, called the best- k rule grouping algorithm. The goal is to compute the best k value under which the performance of the system can be optimized.

The best- k Algorithm

input: EFG, θ , μ , σ , λ_1 , D , and p_o .

output: k and the k -way partition vector ($B[1..N]$).

data structure: the adjacency matrix $A[1..N, 1..N]$.

1. determine N , and $A[1..N, 1..N]$ from the input EFG.
2. set $k = 1$.
3. while $k \leq N$ do

- (a) parameterize r , p , m , and λ based on the output from the KL algorithm: keep the resulting k -way partition.
 - (b) compute the performance metric of the system using Equation (1).
 - (c) if the system performance is the greatest so far, record this value, k , and the k -way partition in the variable $B[1..N]$.
 - (d) $k = k + 1$.
4. output best k and $B[1..N]$.

This algorithm utilizes the KL algorithm [KER 70] for grouping rules such that the total external cost on edge cuts (which results in the k -way partition) is minimized, thereby maximizing r and the system performance for a specified k value. The algorithm runs about 10 seconds for a rule system of size 10 (e.g., AGR) and about five minutes for a rule size of 100 (e.g., MMU).

5 Case Studies

In the following, the best- k rule grouping algorithm is applied to the AGR knowledge base system consisting of seven rules (therefore $N = 7$) and the MMU system consisting of 92 rules. The objective is to identify the best k value and its associated k -way partition which can maximize the effective production rate (i.e., system performance) as a function of λ_1 . For the AGR system, we also compare the theoretical prediction with empirical results to demonstrate the feasibility of our model and algorithm.

The processor on which these knowledge base systems run is an 80386 and the following performance data are observed (e.g., see [WPI 91]):

- $\mu \approx 700 \text{ sec}^{-1}$;
- $\theta \approx 500 \text{ sec}^{-1}$;
- $\sigma \approx 10 \text{ sec}^{-1}$;
- $\lambda_1 \approx 2000 \text{ sec}^{-1}$;
- $D \approx 20$ milliseconds, the average disk service time required for performing a disk read/write operation.

k	r	m	p	λ^* (data-driven)	λ^* (non data-driven)
1	1.00	0	0.00	500	286
2	0.46	1	0.80	1076	571
3	0.32	1	0.80	1555	857
4	0.21	2	0.96	2000	1143
5	0.14	3	0.99	2000	1429
6	0.07	3	0.99	2000	1714
7	0.00	4	1.00	2000	2000

Table 1: r , m , p and λ As a Function of k ($\lambda_1 = 2000$).

$$\bullet \alpha \approx \frac{1}{\frac{1}{2} + \frac{1}{4} + D} = 35.$$

5.1 AGR Knowledge Base System

5.1.1 Theoretical Prediction

By applying the KL algorithm on the EFG of the AGR rule set, it is observed that r , m , p , and λ (under data-driven as well as non data-driven) are related to k ($1 \leq k \leq 7$) as shown in Table 1 when $\lambda_1 = 2000$.

Figure 3 shows that for the AGR knowledge base when the data-driven scheme is used to process arriving facts, the best value of k depends on the magnitude of λ_1 (which reflects the degree of the rule-level overhead). The Figure shows that, if λ_1 is in the same order of magnitude as θ or μ (which occurs when each rule contains many instructions), then the best partition favors a high k . Conversely, if λ_1 is an order of magnitude higher than θ or μ (which occurs when each rule contains only a few instructions), then the best partition favors a low k . The physical interpretation is as follows: in the latter case the time required for the AF to schedule ACOs is an order of magnitude longer than that required for an ACO to process an arriving fact, select a rule, and fire the rule. Consequently, the system performance is improved by shifting the time-consuming scheduling work from the AF to within an ACO. An important observation from this figure is that, for each λ_1 value, there exists a k value under which the system performance is maximized. Since it is estimated $\lambda_1 \approx 2000$ for the AGR knowledge base system, the theoretical prediction suggests that the best partition should group all rules together when the data driven scheme is used.

Figure 4 shows the theoretical prediction for the AGR system when the non data-driven scheme

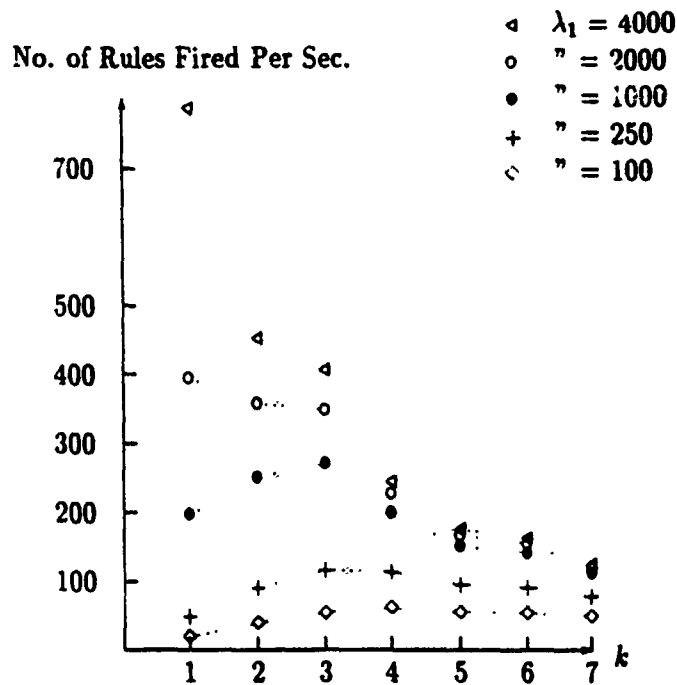


Figure 3: Theoretical Prediction for AGR: the Data-Driven Scheme.

is used to process arriving facts. It predicts the same trend as that of the data-driven scheme except that it shows that a higher value of λ_1 is required for the best partition to favor a low k value since more rules have to be inspected when a new fact arrives. Comparing Figure 3 with Figure 4, we observe the following: (a) the performance of the data-driven mechanism is always better than the non data-driven mechanism, and (b) for the same λ_1 value, the best k value selected under these two schemes may be different due to a different degree of rule-level overhead, e.g., when $\lambda_1 = 2000$, Figure 4 predicts that $k = 3$ is the best selection for the non data-driven scheme, while Figure 3 predicts that $k = 1$ is the best selection for the data-driven scheme.

5.1.2 Comparison To Experimental Results

Figure 5 compares the theoretical prediction with the experimental results for the AGR system where the y-coordinate represents the performance ratio between the one-group case and the k -group case with the x-coordinate varying k from one to seven. This experimental result is obtained by encoding the AGR rules in the form of Activation Framework Language (AFL) rule groups based on the output generated by the best- k rule grouping algorithm and having the AFL translator [GRE 91C] automatically translate them into groups of C program modules incorporating the data driven

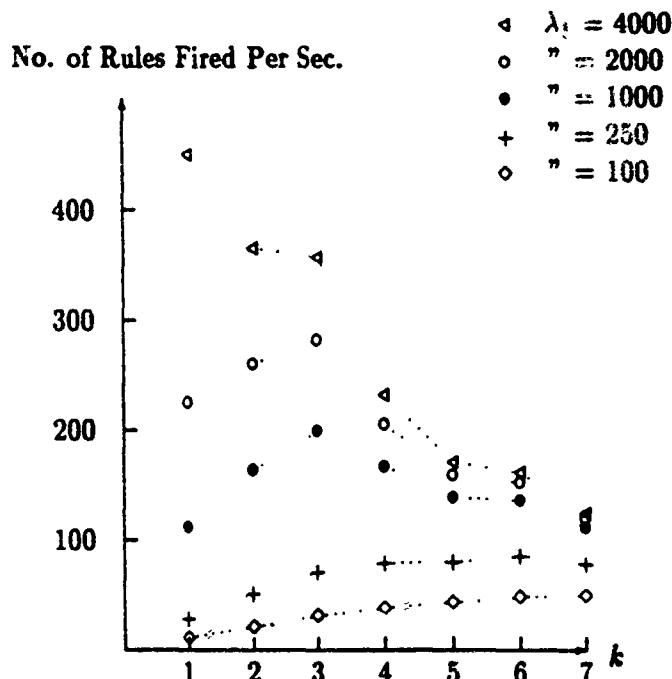


Figure 4: Theoretical Prediction for AGR: the Non Data-Driven Scheme.

mechanism as described in [GRE 91B]. These C program modules are then subsequently linked with AFO test modules and the Activation Framework For C Language (AFC) run-time libraries into one executable program. To ensure that no injected data are lost, the times at which data are injected into the system are manually adjusted (separately for each k value) in the input file (input.txt) such that the execution order of rules is repeated, e.g., rule 2 first, followed by rule 6 and then by rule 3, etc. The time required for AGR to output an expected sequence of logging records in the output file (output.txt) is measured separately for each k value and is used as the performance standard with which the theoretical prediction is compared. It was observed that, including the disk operation overhead required to inject and log the data, 7.19 seconds are needed for seven groups, 3.3 seconds are needed for two groups and 2.3 seconds are needed for one group for the expected execution order to be observed.

Figure 5 shows that, for the data-driven scheme, the theoretical prediction correlates well with the experimental results. The result that one group is the best selection is not surprising because the AGR system consists of only seven rules and therefore the rule-level overhead is not significant when compared with the AFO-level overhead, especially when the data driven mechanism is utilized. In the following we will show that for a system with moderate size, e.g., the MMU system with 92 rules, one group is not necessarily the best selection because as the the number of rules increases,

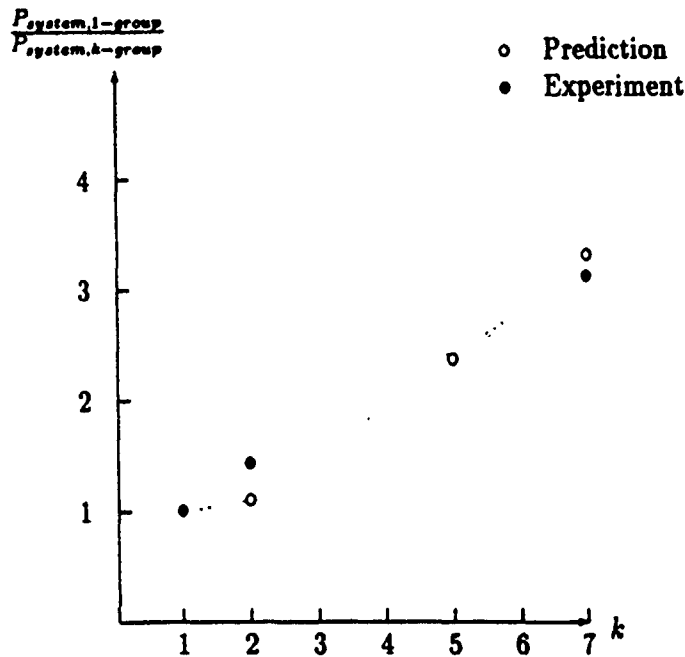


Figure 5: Theoretical Prediction vs. Experimental Result for AGR System.

the rule-level overhead also increases significantly since all rules are put into one group.

5.2 The MMU Knowledge Base

Figure 6 shows the theoretical prediction for the MMU knowledge base system (for data-driven scheme). The result suggests that four groups should be the best partition for $\lambda_1 = 2000$ although the result remains to be validated in the future.

6 Summary

In the report we have developed a theoretical basis with which the performance of rule grouping algorithms can be evaluated. The theoretical basis derives from the concept of a finite state machine by considering the execution of the rule system running on the AF architecture as a progression through states representing the rule-level overhead (the ACO state) and the AFO-level overhead (the remaining states), respectively. Based on the model, we derived the system performance equation. A best- k rule grouping algorithm was developed utilizing (a) the KL graph-partitioning heuristic for k -way partitioning the rule system and (b) the system equation for determining the best k value under which the system performance can be maximized. The results showed that the

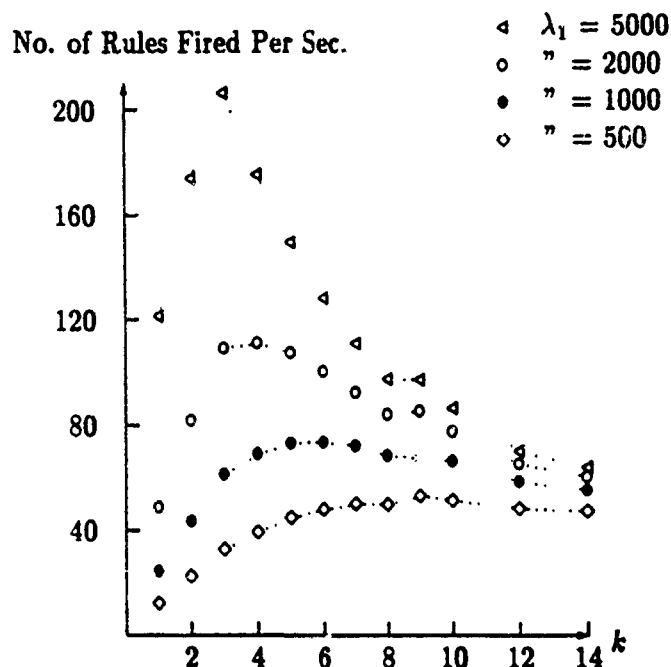


Figure 6: Theoretical Prediction for MMU: the Data-Driven Scheme.

best k -way partition is dictated by the balance between the rule-level and AFO-level overheads: if the rule-level overhead dominates the AFO-level overhead (e.g., when there are a lot of rules, or a lot of instructions per rule and/or the non data-driven mechanism is used), a high group number is favored; conversely, if the AFO-level overhead dominates the rule-level overhead (e.g., each rule is simple) a low group number is favored. In general, the best k -way partition has to be determined from the performance equation because one overhead may not always dominate the other. Finally, to demonstrate the feasibility of our approach, we compared the theoretical prediction from the theory and algorithm with empirical results for the AGR knowledge base system and a good correlation was observed.

Some possible future research directions include (a) further validating the model for the single processor AF architecture by comparing theoretical predictions with experimental results for medium (100 rules) to large (1000 rules) knowledge base systems. (b) investigating whether an unbalanced k -way partition can perform better than a balanced k -way partition by refining the model to accommodate groups of different size, (c) incorporating the notion of group importance level to more accurately estimate computable parameters, (d) refining the modeling of the data-driven mechanism by including the overhead required for updating the rule importance levels and selecting the most important rule to fire within a group, and (e) developing and validating a the-

oretical basis for evaluating the performance of rule grouping algorithms for rule systems running on a distributed multiprocessor AF architecture based on the experience learned from the single processor AF architecture.

References

- [BAR 82] Barnes, E.R., "An algorithm for partitioning the nodes of a graph" *SIAM Journal on Algebraic and Discrete Methods*, Vol. 3, No. 4, Dec. 1982. pp. 541-550.
- [FOR 82] Forgy, C.L., "Rete: a fast algorithm for the many pattern/many object pattern match problem," *Artificial Intelligence*, 1982, pp.17-37.
- [GRE 87] P.E. Green, "AF: A framework for real-time distributed cooperative problem solving," *Distributed Artificial Intelligence*, edited by Michael N. Huhns, Morgan Kaufmann Publishers, Inc., 1987, pp. 153-176.
- [GRE 91A] P.E. Green, J. Duckworth, L. Becker, and S. Cotterill, "Maintenance system for AI knowledge bases, Phase I - final report, Design of the AKRAM system", contract number F33615-90-C-1470, 1991.
- [GRE 91B] P.E. Green, "A data driven mechanism for the execution of production rules in real-time computer based systems", *The Real-Time Intelligent Systems Cooperation*, 1991.
- [GRE 91C] *AFL users manual*, Version 1.0, *The Real-Time Intelligent Systems Cooperation*, 1991.
- [JAC 90] Jacob, R.J.K., and Froscher, J.N., "A software engineering methodology for rule-based systems," *IEEE Trans. Knowledge and Data Eng.*, Vol. 2, No. 2, June 1990. *Computer-Aided Design*, Vol. 21, No. 10, Dec. 1989, pp. 611-618.
- [KER 70] Kernighan, B.W., and Lin, S., "An efficient heuristic procedure for partitioning graphs," *Bell Sys. Tech. J.*, Vol. 49, 1970, pp. 291-307.
- [LEE 89] Lee, C-H, Park C-I, and Kim M., "Efficient algorithm for graph-partitioning problem using a problem transformation method,"
- [MEH 91] Mala Mehrotra, "Rule groupings: a software engineering approach towards verification of expert system," *NASA Contract NAS1-18585*, Final Report, Feb. 1991.
- [ROS 89] S.M. Ross, *Introduction to Probability Models*, New York, Academic Press, 4th Edition, 1989.
- [RTI 91] *The Real-Time Intelligent Systems, Activation Framework Operating System Environment, Users Manual*, AFC Version 2.5, 1991.
- [WPI 91] Worcester Polytechnic Institute, *Knowledge Representation into Ada Methodology*, Project Review Meeting, May, 1991.

HIGH TEMPERATURE OHMIC CONTACTS FOR GaAs

Final Report

by

David W. Moss
Graduate Student of Physics
Wilkes University, Wilkes-Barre, PA. 18766

on

Work done at the Research Division of Solid State Electronic
Technology Directorate, Wright Laboratory, Wright-Patterson
Air Force Base, Ohio 45433-6543

Dates: May 18, 1991 - July 27, 1991

under

1991 Summer Research Program for Faculty and Graduate
Students of the Air Force Office of Scientific Research

Contract # F49620-90-C-0076

Abstract

Fabrication of electrical contacts to n+ and p+ GaAs were studied using GeMoW as the refractory ohmic metal and TiPtAu, TiPtAlTiPtAu, and Al were used as the different Schottky metals. The transmission line model (TLM) was used to determine the contact resistance at room temperature. The contacts were annealed with As and InAs over pressure. The samples were annealed at increasing temperatures, then the contact resistance was measured after each anneal. Plots of average contact resistance verses annealing temperature were created to help understand the relation between the two parameters. A GeMoW/TiPtAu contact on p+ GaAs had the lowest average minimum contact resistance of 0.14 Ω -mm after 550 °Celsius anneal and remained as an ohmic contact until 650 degrees. A GeMoW/Al contact on p+ GaAs remained ohmic from 300 ° C to 550 ° C with an average minimum contact resistance of less than 0.4 Ω -mm. The goal of this work was to increase the operating temperature of GaAs based HBTs and MESFET devices by eliminating ohmic contact as an impediment.

1. INTRODUCTION

As the dimensions of semiconductor devices continue to decrease, the importance of low resistance contacts to these devices becomes an increasingly important limiting factor. Even with the present technology the contact resistance in GaAs devices may be half of the total parasitic resistance.¹ This shows the vital importance of developing a reliable metal system to be used as electrical contacts to GaAs and other III-V semiconductors. With the advances in the crystal growth, device concepts, and fabrication of electronic and optical devices, the importance to increase the reliability for ohmic contacts at high temperatures is a growing concern. For high temperature applications in microelectronics, the solid state devices are required to function reliably in hot ambient ($>300^{\circ}\text{C}$).² The excellent material properties and the current advances in technology have made GaAs a leading contender among the III-V semiconductors for high temperature applications.³

2. OHMIC CONTACTS

There are many definitions for an ohmic contact, which include the following:

- (1) a perfect source and sink of both carrier types having no tendency to inject or collect either electrons or holes;
- (2) a source of carriers with an internal resistance R_c which is totally negligible compared with the semiconductor resistance;

(3) a source of carriers with a non-negligible internal resistance R_C , but one which obeys Ohm's law for current densities of interest.⁴

Definitions two and three will be used throughout this work. Definition one is impractical in experiment because no contacts ever actually fulfill this definition.

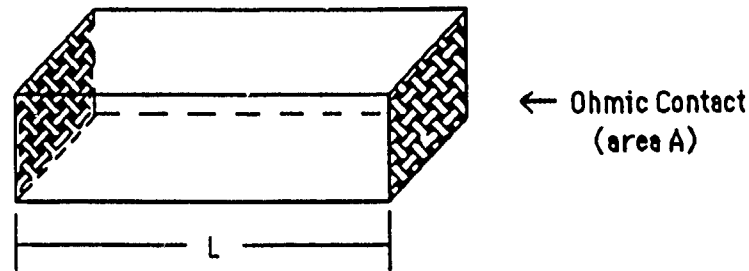


Fig. 1

A slab of semiconductor with ohmic contacts at each end.

First, it is important to understand the definition of contact resistance. There is a resistance associated with a metal to semiconductor junction. Figure 1 shows a slab of semiconductor with length L and cross-sectional area A . An ohmic contact makes a junction with the semiconductor at each end of the slab. The resistance of the semiconductor material from end to end is:

$$R = \rho L/A \quad [1]$$

where ρ is the resistivity of the semiconductor. There is also a contact resistance due to each ohmic contact, R_C , therefore the total

contact resistance is:

$$R = 2R_C + \rho L/A \quad [2]$$

A fundamental quantity that defines the quality of a contact is the specific contact resistance. The specific contact resistance, r_C , is the contact resistance per a unit area for the current flow perpendicular to the contact and has units of $\Omega\text{-mm}^2$. For figure 1, the specific contact resistance is:

$$r_C = R_C A \quad [3]$$

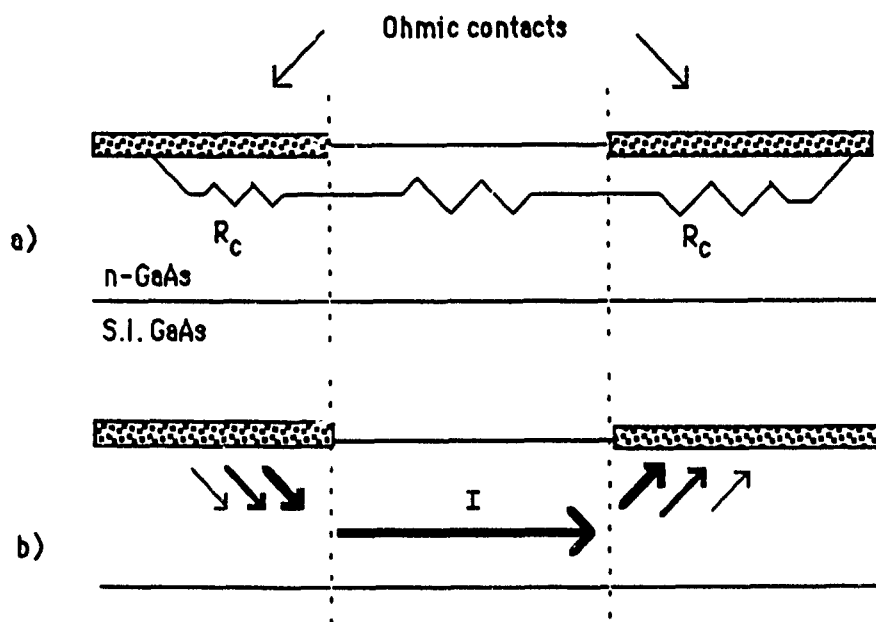


Fig. 2

The planar configuration of ohmic contacts.

Many of the modern GaAs devices have a planar configuration which

increases the complexity of the measurement. (see fig. 2) The ohmic contact is formed on a thin, conductive layer having a semi-insulating layer beneath it. The final direction of current flow is parallel to the plane of the metal instead of perpendicular to it. In this situation the contact resistance is defined as the resistance between the metal and an imaginary plane that is at the edge of and perpendicular to the contact. Most of the current flows into the contact relatively near the edge. The magnitude of current remaining in the semiconductor decreases exponentially from the contact edge.⁵ This contact resistance is a function of both the specific contact resistance and the characteristics of the conducting layer on which the contact is formed. This method of defining contact resistance is found using the transmission line model (TLM), which is covered in the next section.

3. TRANSMISSION LINE MODEL

The transmission line model (TLM) was originally proposed by Shockley in 1964.⁶ This model offers a method for determining the contact resistance for planar ohmic contacts. The TLM determines the ohmic contact parameters in the following way; resistance measurements R_1 and R_2 are made between adjacent contacts as shown in Figure 3.

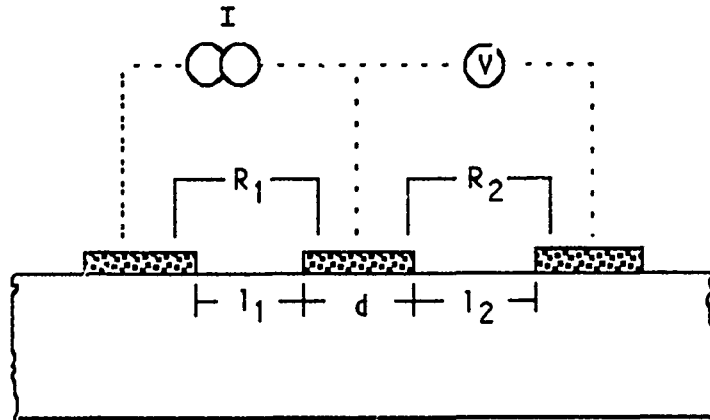


Fig. 3
Experimental measurements for obtaining total contact resistance.

The total resistance, R_T , between any two adjacent contacts is:

$$R_T = 2R_C + (R_{sh} \cdot l)/W \quad [4]$$

where R_{sh} is the sheet resistance of the semiconductor layer outside the contact region, l is the distance between the contact, and W is the width of the contact. The contact resistance, R_C , can be shown to be equal to:⁷

$$R_C = [(R_{sk} \cdot L_T)/W] \cdot \coth(d/L_T) \quad [5]$$

where R_{sk} is the sheet resistance of the layer directly under the contact, d is the length of the contact, and the transfer length can be defined as :

$$L_T = [(P/R_{sk})]^{1/2} \quad [6]$$

$$\text{when } d \geq 2L_t, \quad R_T = [2R_{sk} \cdot L_T + R_{sh} \cdot l]/W \quad [7]$$

The relationship of equation 7 can be seen in figure 4. This figure is a plot of total contact to contact resistance as a function of l , the distance between adjacent contacts, to obtain transfer length and contact resistance values.

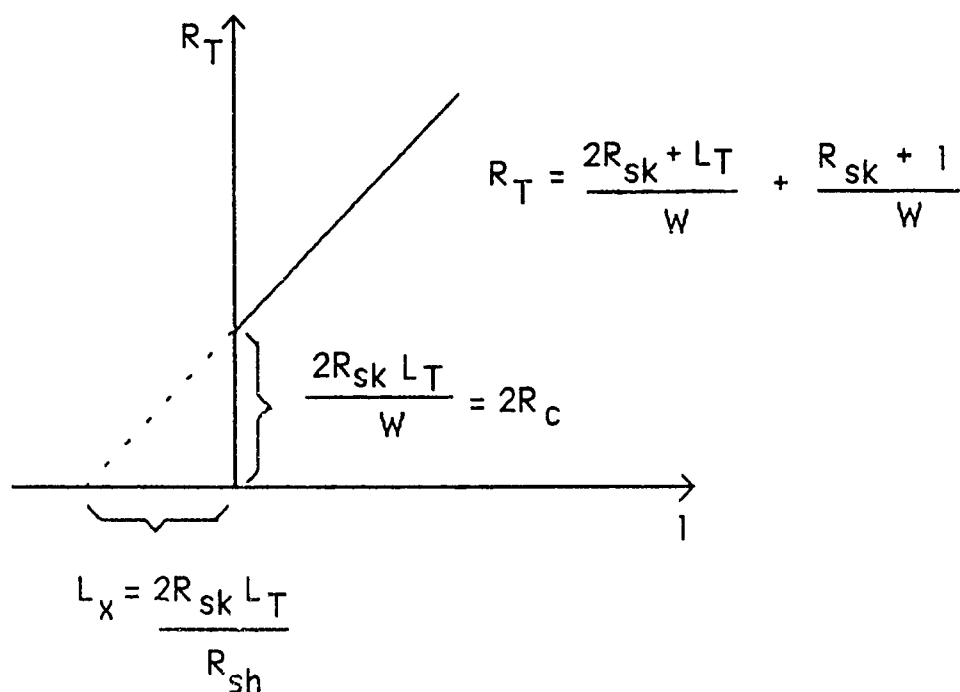


Fig. 4

Plot of total contact to contact resistance as a function of l to obtain transfer length and contact resistance values.

4. EXPERIMENTAL PROCEDURE

This work was done at Wright Patterson Air Force Base's Research

Division of Solid State Electronic Technology Directorate in Dayton, Ohio.
All work was done in the laboratories class 100 clean room, excluding the annealing of the samples.

Both n+ and p+ samples were grown with different amounts of doping, in order to have a variety of samples to process. Samples were grown using the Varian GEN II molecular beam epitaxial growth system. Samples were requested with specific types and amounts of dopants required. The objective was to develop refractory metal systems for high temperature ohmic contacts and Schottky barriers for both n+ and p+ GaAs. The following is a list of the samples used in this experiment:

<u>Sample #</u>	<u>Type</u>	<u>Layer thickness</u>	<u>Doping concentration</u>
1129	n+ GaAs	1 μm	5×10^{18}
1130	p+ GaAs	3000 \AA	1×10^{19}
1131	n+ GaAs	3000 \AA	5×10^{18}
1132	n+ GaAs	3000 \AA	5×10^{18}
1133	p+ GaAs	1 μm	8×10^{18}
1137	n+ GaAs	3000 \AA	5×10^{18}
	InGaAs	100 \AA	5×10^{18}
1138	n+ GaAs	3000 \AA	5×10^{18}
	InGaAs	50 \AA	5×10^{18}

After the samples were grown, some were tested by Hall measurement to determine the validity of the dopant concentrations. These results will be listed in the data section of this report. The procedure following the growth of the samples was:

1. Cleave the wafer into four equal pieces, labeled A, B, C, D.
2. Use the MIMIC 4E mask to perform the mesa isolation lithography.
3. Etch the mesas, using a 1:1:8 solution of HF:H₂O₂:H₂O, until the mesas are isolated.
4. Strip the photoresist and measure the mesa height.
5. Use the 4E mask to perform S/D lithography using double PMGI/PR photoresist process.
6. Remove the PR, clean with acetone and isopropyl alcohol and descum the surface.
7. Prepare the sample for metal deposition.

After the lithography processes, the metal deposition was done. The contact metals were either deposited by evaporation or sputtering. All the materials, except tungsten, were evaporated and the tungsten was deposited by sputtering. After the deposition, the excess metal was lifted off to have the metal only on the contacts and not over the entire wafer. The typical layer sequence and thicknesses for the refractory ohmic contacts were 100Å Ge/200Å Mo/ 1000Å W. These values were typical and changed in later samples to see how the thickness changes correlated to improve the contact resistance of the sample. The Schottky/second level metals used were limited to three different systems; a) Ti/Pt/Au b) Al and c) Ti/Al/Ti/Pt/Au.

After the samples were processed, the samples were taken to the DC testing area. The contact resistance of the samples were measured using the TLM method. A mainframe computer system was used to help collect the data and to allow transmission of the data from inside the clean room to the surrounding offices in the building. The Keithly Yield Max 450 was the process monitor system used for the micro-vax system in the DC

testing area. Each subsite on the sample contained fourteen TLM patterns, seven gated patterns, and seven ungated patterns (see figure 5). Depending on where the sample was cleaved there could be as many as fifteen subsites on the sample. The testing was done on a Electroglas model 200 LX which allowed for the sample to be maneuvered under the probe card efficiently. The probe card was interfaced to the Keithly so that all information could be recorded and the input current maintained at a constant level. Originally, the samples were aligned by eye under a microscope to the various testing sites on the quarter wafer. Then each TLM site was individually tested. This was a very tedious approach and limited the number of sites that could be tested in a reasonable amount of time. A program which automated the testing procedure was introduced. This program enabled all the available sites on the sample to be tested and the data to be recorded in the same time it took to previously measure ten TLM sites. The initial measurement of the contact resistance was taken at room temperature before annealing the sample. After the room temperature data was collected the samples were annealed. The high temperature annealing furnace had independent temperature control of As and InAs for over-pressure and doping control. The over-pressure of As and InAs limited diffusion from the sample during the anneal. The samples were annealed to 300 ° Celsius for ten minutes and then, the contact resistance remeasured. This cycle continued for each sample, annealing at a temperature 50 degrees higher and contact resistance measured, until the samples contact resistance became non ohmic. The criteria for non-ohmic behavior was when the correlation of the data in the plot of the samples total contact resistance as a function of contact spacing went below 0.90 , using a least squares fit for the data. This least squares fit of the data was included as a subroutine in the program used when testing

the samples.

5. RESULTS

The results of the Hall measurements taken on samples 1130 and 1133 agreed with the desired carrier concentrations for the samples. The Hall measurements were taken at room temperature after 550 ° anneal and after 770 ° anneal. The results for samples 1130 and 1133 averaged over current and magnetic field were:

sample: 1130

<u>anneal temperature</u>	$\rho(\Omega\text{-cm})$	$\mu(\text{cm}^2/\text{V}\text{-sec})$	<u>conc.</u> (cm^{-3})
room	9.20×10^{-4}	62.1	3.0×10^{19}
550 °	9.19×10^{-4}	61.0	3.0×10^{19}
770 °	3.17×10^{-3}	59.0	3.1×10^{19}

sample: 1133

<u>anneal temperature</u>	$\rho(\Omega\text{-cm})$	$\mu(\text{cm}^2/\text{V}\text{-sec})$	<u>conc.</u> (cm^{-3})
room	7.86×10^{-3}	916	8.7×10^{18}
550 °	7.84×10^{-3}	901	8.8×10^{18}
770 °	7.95×10^{-3}	865	9.0×10^{18}

Sample 1130 showed the best results for lowest contact resistance for p+ GaAs. This sample used Be, as the dopant, during MBE growth. TiPtAu and Al were used as the Schottky barrier and second level metal. The contact resistance varied from 0.13 - 0.29 $\Omega\text{-mm}$ from 300 to 650 ° C for TiPtAu and from 0.23 - 0.27 $\Omega\text{-mm}$ between 300 to 500 ° C for Al. Sample 1132 showed the best results for a n+ sample. The contact

resistances varied from 0.27 - 0.30 Ω -mm from 350 to 500 ° C for TiPtAu and from 0.30 - 0.34 Ω -mm between 350 to 500 ° C for Al. The contact resistance versus anneal temperature plots are shown for these two samples along with the plots for samples 1137 and 1138, which also showed the most promising values for the contact resistance.

6. CONCLUSION

The development of the high-temperature ohmic contacts using GeMoW as the refractory metal layer is a promising area of research. By increasing the operating temperature of these contacts, devices will be able to be used at elevated temperatures without having the parasitic effect of many high temperature contact systems. The best system for p+ GaAs was found to be GeMoW/TiPtAu. The lowest contact resistance was found to be 0.13 Ω -mm at 550 ° C. The best system for n+ GaAs was GeMoW/TiPtAu with a contact resistance of 0.30 Ω -mm at 500 ° C. The system of GeMoW/TiPtAu can be used for both n+ and p+ GaAs at temperatures of 500 ° C. It is imperative to have a very high level of doping ($> 5 \times 10^{18}$) at the surface layer, for GeMoW to be successful ohmic contact. There are still some problem areas that need to be approached in order for GeMoW to reach its full potential. One of these problems which needs to be faced in the future is the outdiffusion of Be at higher temperatures in p+ GaAs. One solution is to dope the p+ samples with a different dopant. The most promising choice would be carbon doping. Another problem area is that S.I. GaAs substrates become conductive at approximately 245 ° C.⁸ One possible solution is to peel processed active layers from the substrate and Van der Waals bond them to

another substrate (diamond, Si, ceramic).

GeMoW as the refractory ohmic contact is a promising system for high temperature use. The fact that it can presently be used for both n+ and p+ GaAs at 500 ° shows signs that it's operating temperature can be increased. Future research is needed to increase this practical operating temperature to its fullest potential. Fortunately the Electronics Laboratory Research Division at Wright Patterson Air Force Base is presently continuing its research on ohmic contacts using GeMoW as its refractory metal system.

ACKNOWLEDGEMENTS

The author would like to acknowledge the warm hospitality of the Research Division of the Solid State Electronics Directorate of Wright Laboratory at the Wright Patterson Air Force Base. Their cooperation and guidance made this stay a very valuable and enjoyable experience. In particular, Gary McCoy, Chern Huang, and Ken Nakano were excellent hosts and created a perfect atmosphere for research. Captain Ken Merkel provided wonderful leadership and a extremely helpful mentor during my ten week research period. I am indebted to Lt. Ross Detmer for the countless hours of help with the programming and guidance for the TLM measurements. I would also like to thank Gerry Robinson, Lt. Jerry Trombley and Chris Bozada for providing valuable help throughout my stay for which I am grateful.

REFERENCES

1. C. Dubon-Chevalier, P. Blanconnier, and C. Besombes, J. Electrochem. Soc., **137**, 1515(1990)
2. R. P. Gupta, W. S. Khokle, J. Wuerfl, and H. L. Hartnagel, J. Electrochem. Soc., **137**, 631(1990)
3. R. P. Gupta, W. S. Khokle, J. Wuerfl, and H. L. Hartnagel, J.

- Electrochem. Soc., **137**, 631(1990)
4. D. Look, *Electrical Characterization of GaAs Materials and Devices*, John Wiley & Sons, New York, 1989, p. 22.
 5. R. Williams, *Modern GaAs Processing Techniques*, Norwood, MA., 1990, p. 213.
 6. G. K. Reeves and H. B. Harrison, *Electron Device Letters*, **3**, 111(1982)
 7. G. K. Reeves and H. B. Harrison, *Electron Device Letters*, **3**, 112(1982)
 8. G. Robinson, WL/ELRA, Wright Patterson Air Force Base, Dayton, OH., unpublished report (1991)

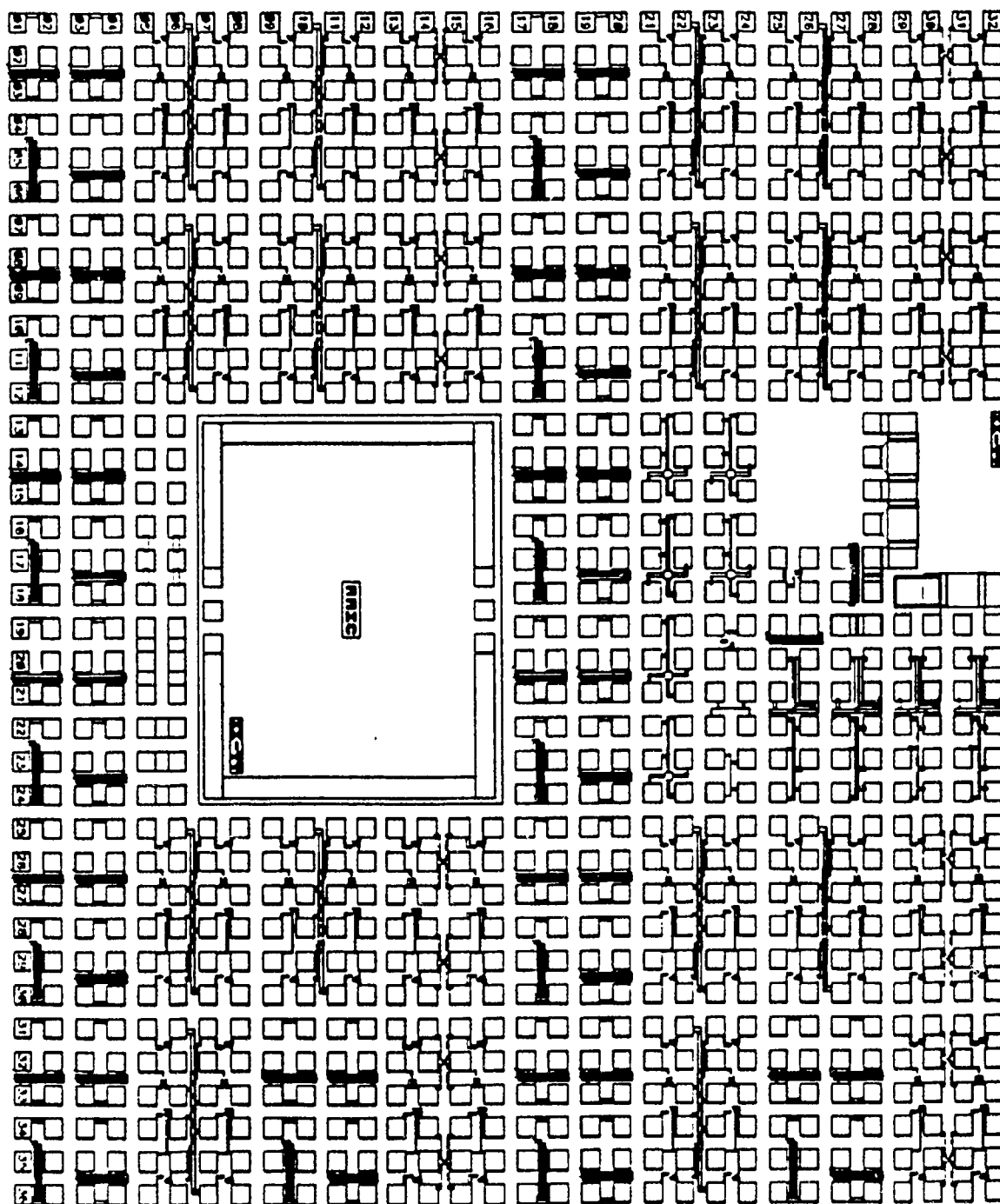


Fig. 4

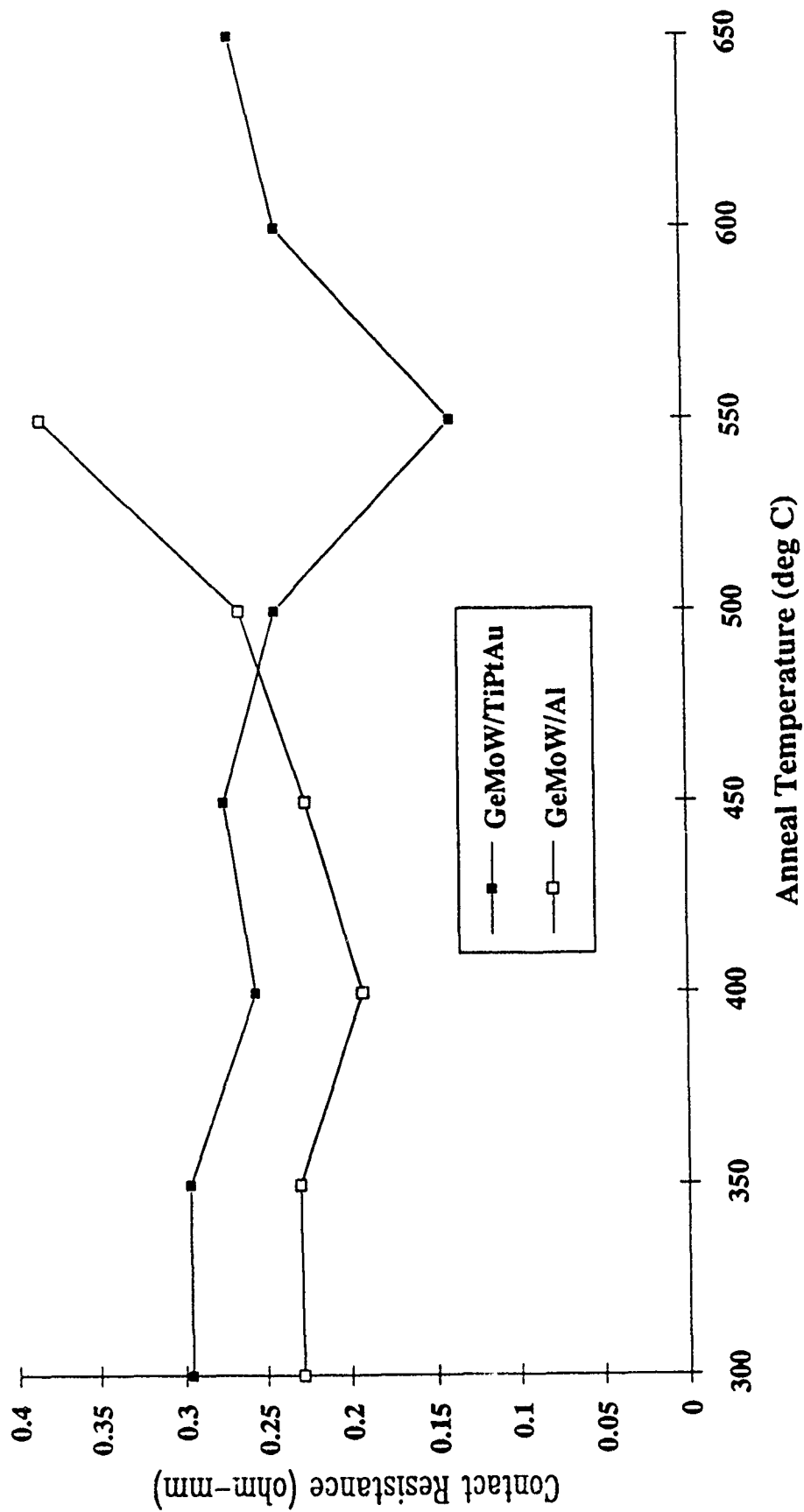
Gated TLM

UnGated TLM

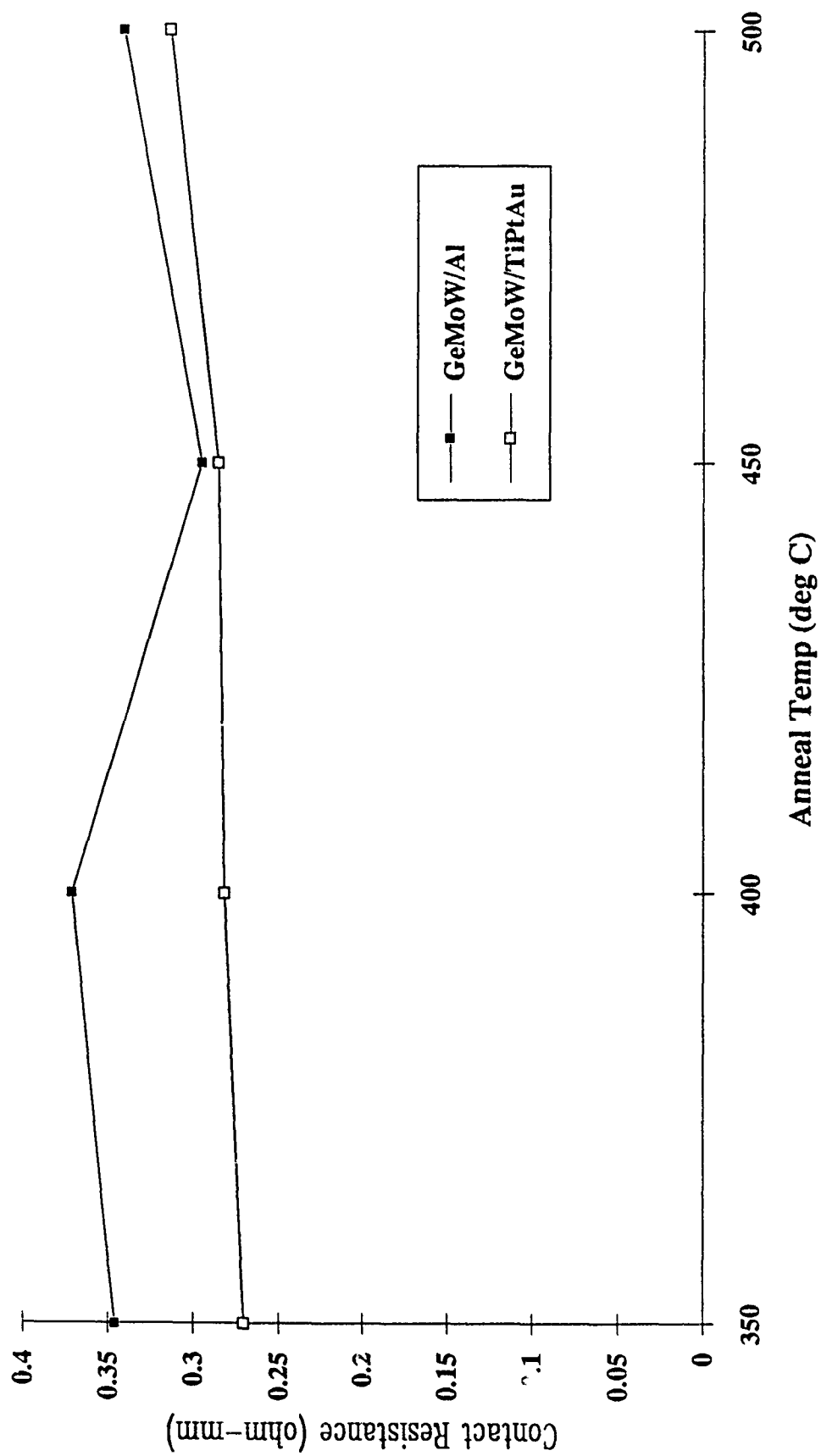
Gated TLM

Inverted TLM

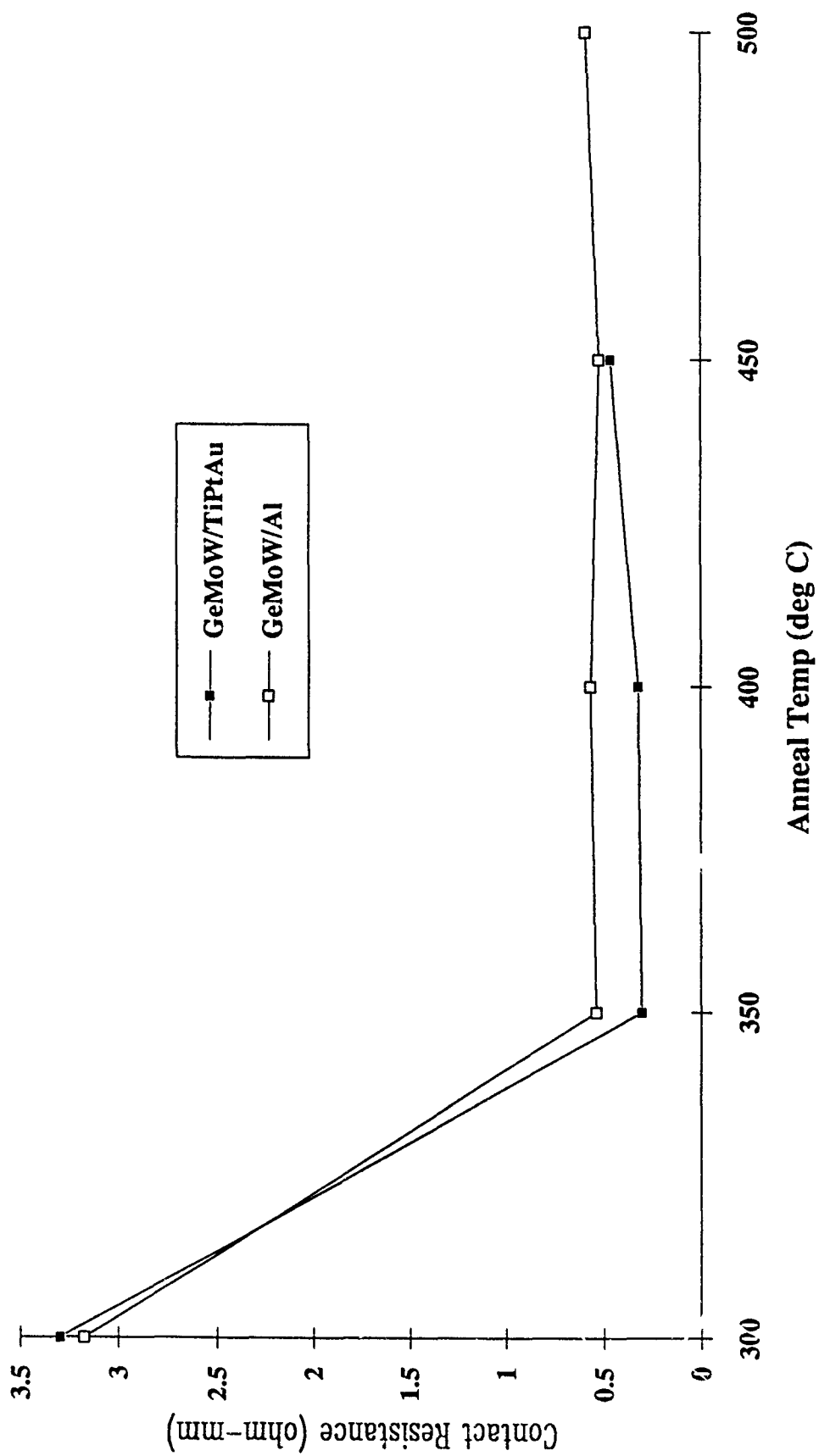
Average Contact Resistance vs Annealing Temp, G21130A (p+ GaAs) , GeMoW/Al & GeMoW/TiPtAu, Ungated TLM



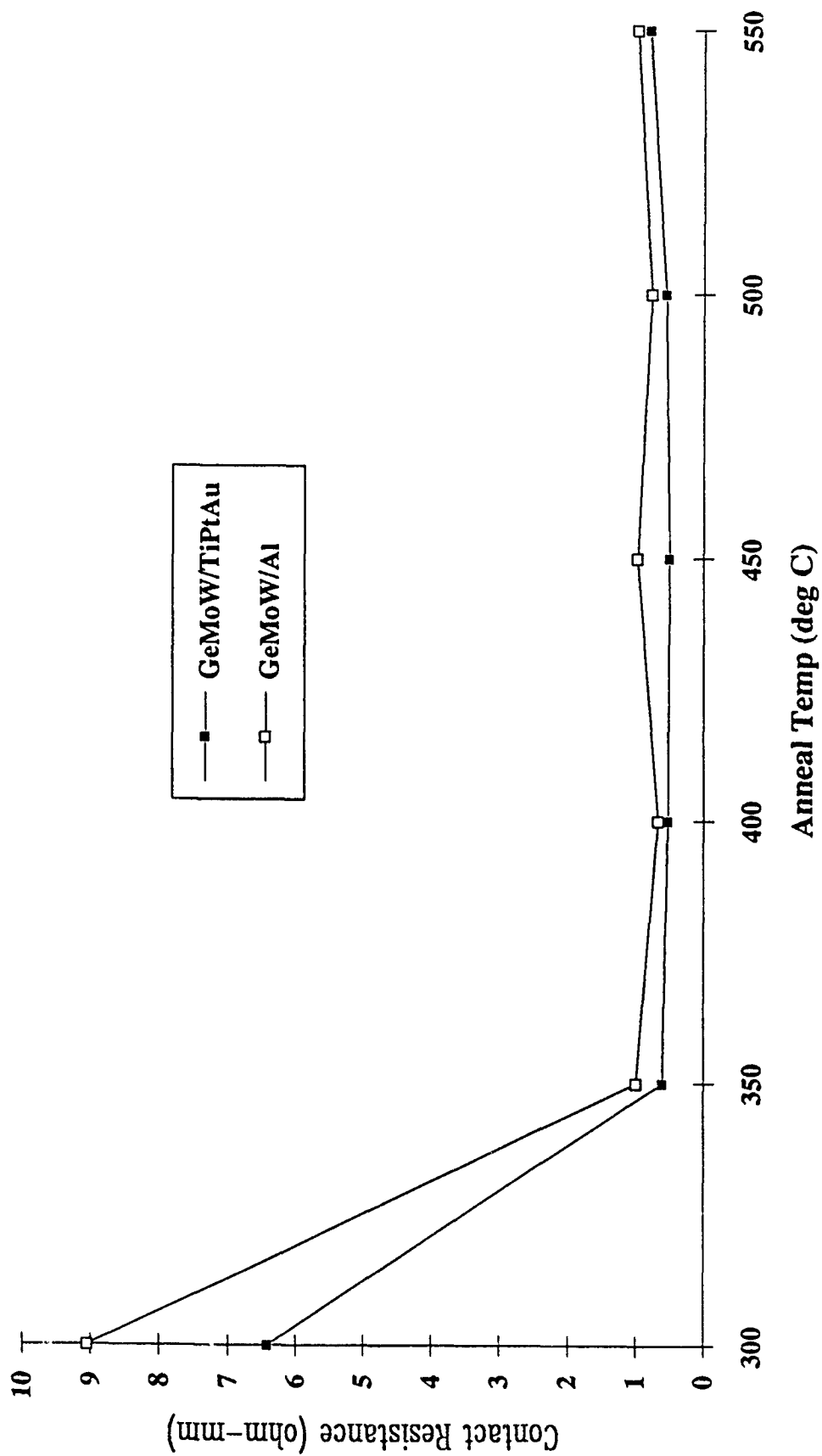
Contact Resistance vs Anneal Temp, G21132 (n+ GaAs), GeMoW, Ungated TLM



Average Contact Resistance vs Anneal Temp, G21137 (n+ InGaAs/n+ GaAs), Ungated TLM



Average Contact Resistance vs Anneal Temp, G21138 (n+ InGaAs/n+ GaAs), Ungated TLM



1991 USAF-RDL SUMMER FACULTY RESEARCH PROGRAM/
GRADUATE STUDENT RESEARCH PROGRAM
Sponsored by the
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Conducted by the
THE RESEARCH & DEVELOPMENT LABORATORIES
FINAL REPORT
ACOUSTIC EMISSION BEHAVIOR AND DAMAGE ACCUMULATION
MONITORING OF CORD-RUBBER COMPOSITE MATERIALS

Prepared by:	Pat K. Hippo
Academic Rank:	Graduate Student
Department and University	Engineering Science and Mechanics The Pennsylvania State University
Research Location	Wright-Patterson Air Force Base Wright Research Laboratory Flight Dynamics Directorate Vehicle Subsystems Division Aircraft Launch and Recovery Branch
USAF Researcher:	John P. Medzorian
Date:	03 September 1991

ABSTRACT

Relying on coupon representations simulating the aircraft tire carcass region, this study investigated the key parameters that control damage accumulation in cord-rubber composite materials. Acoustic emission (AE) behavior of these specimens was monitored as well and the relationship between damage accumulation and AE was assessed for static and cyclic loadings. The results showed that the overriding factor controlling damage accumulation was stress amplitude rather than mean stress. In addition, strain dependence of AE under static loading was clearly established. The results were not as clear under cyclic loading conditions as the relationship seems dependent on selective minimum loads. However, AE potential as a predictor of imminent failure of aircraft tires still has some merit as refined procedures and improved instrumentation may eventually prove.

INTRODUCTION

Currently, the operational life of bias-ply aircraft tires used by the U.S. Air Force is qualified by expensive dynamometer testing and analysis that specify the number of takeoff and landing cycles a tire may endure before replacement. An initiative is underway to realign this procedure where emphasis is placed on more cost-effective testing supplemented with computer modelling and finite element analysis. By using this strategy, the initial goal is to identify the parameter(s) that control initiation threshold and damage accumulation in cord-rubber composite materials. Eventually, it is hoped a fatigue law can be formulated that will serve to predict tire carcass durability on a reliable basis. Both efforts could extend tire replacement intervals which ultimately would prove more cost-effective than existing criterion that now govern tire replacement.

The focus of this study addresses the first goal i.e., parameters controlling damage accumulation with special inclusion of establishing AE as a viable experimental technique for real-time monitoring of damage accumulation. Last year (Summer 1990), a preliminary study investigated the deformation and fracture behavior of a variety of cord-rubber composite materials under different loading regimens. One conclusion was that damage accumulation in coupon specimens was accompanied by a steady increase of temperature and local strain. In addition, these specimens were found to exhibit extensive interply shear deformation which eventually led to delamination type failure.

Interestingly, the effects of temperature, stress and strain can induce acoustic events that are associated with the sudden release of discrete bursts of energy. Based on this premise, appropriate monitoring of AE might be

expected to lend insight into damage accumulation and the mechanisms controlling fatigue lifetimes of these materials. Most importantly is the fact that numerous studies have shown different failure modes generate distinctively different acoustic signatures; therefore its use as an NDE tool for monitoring structural integrity could prove quite successful and valuable, particularly if the acoustic signature of the failure mode is known a-priori. Since the study conducted last year concluded that specimens failed by delamination, there is great potential for using AE as a predictor of imminent tire failure.

Surprisingly, little work has been initiated with cord-rubber composite materials using AE analysis even though it is a technique that has been used for years with other structural materials such as plastic matrix composites and metals to analyze fracture. An AE is a sound wave, or more properly, a stress wave that travels through a material. As noted earlier, the source is a sudden release of energy induced by stress. AE technology involves the detection and conversion of these mechanical stress waves into electrical voltage signals by listening devices called transducers that are mounted on the specimen surface. When a load is applied to a specimen, stress waves initiate particle motion that subsequently exert pressure on the face of the transducer. Signal amplification and electronic processing of these transmitted waves then produces an acoustic signature.

Although described in a simplistic manner here, AE does have some shortcomings that can limit its effectiveness. For example, external sources of noise may interfere with analysis. Often, the proper selection of transducers (i.e., frequency range) and utilization of selective filter and threshold frequency settings on the AE instrument will minimize unwanted

information, but they do not necessarily ensure their complete elimination as some noise is simply an inherent problem (e.g., gripping noise from cyclic loading) and can not be fully removed. In other cases, its application to a particular structure may be difficult. For example, monitoring AE of an aircraft tire in service poses some obvious problems. Technology gaps that now exist would have to be overcome in a cost-effective manner.

OBJECTIVE

The present investigation was undertaken in accordance with continuing efforts to predict aircraft tire durability. Two objectives were set forth; the first was to identify the parameters controlling damage accumulation and structural failure in cord-rubber composite materials while the second was to correlate AE behavior with these processes under both static and dynamic fatigue loadings.

APPROACH

A cord-rubber composite material consisting of 2 + 2 x 0.25 brass-coated steel reinforcement and a proprietary rubber compound matrix was considered for this research. Laminate panels (8" length x 3" width x 0.25" thick) were supplied by the Goodyear Tire & Rubber Co. and were comprised of hand lay-ups with a 4-ply $\pm 23^\circ$ symmetric geometry. Additional ply build-ups were constructed at the ends to protect the material from grip damage that could be experienced during mechanical testing. Coupon specimens approximately 0.75" wide were machined from these panels and polished on a grinding wheel in preparation for the testing. As Figure 1 illustrates, these angle-ply composite coupons were designed to maximize

interply shear deformation occurring in the carcass region of a typical bias-ply tire.

The advantages of coupon testing rest with the easy application to uniaxial tension instruments and the relatively low cost of such testing. Also, the coupon material is nearly identical to the tire carcass footprint region where damage accumulation and catastrophic failure might be expected to occur. For these reasons, coupon testing was selected as the main approach for reaching the objectives.

It can be argued that coupon testing is too simple of an approach for accurately describing the multi-axial loadings a tire actually experiences. Free-edge effects from the cut ends of reinforcing cords also pose some concern as well. However, it is important to realize that individual structural failure processes observed in coupon specimens could be very representative of an aircraft tire. AE could play an important role in justifying coupon testing as a viable approach.

Two loading regimens were considered in this study - they included static tension and cyclic tension. With regard to static loading, displacement rate was kept constant while outputs of displacement/load and AE events, counts, and energy were recorded. In the case of cyclic loading, various combinations of mean stress and stress amplitude were implemented. Table 1 shows the test matrix involving these parameters and the various loading combinations selected. Outputs of displacement/Kcycle, temperature/Kcycle as well as AE events, counts and energy were monitored. Two specimens were tested under each combination and in all cases, a cycling frequency of 5 Hz was employed. For both types of loading, periodic photographs were taken using a camera with a special macro-lens so interply shear strains

could be calculated.

AE monitoring was accomplished with Physical Acoustics Corp. (PAC) LOCAN-AT instrumentation and two R50 piezoelectric transducers. One transducer was connected to a wide bandpass preamplifier having a filter frequency of 100 - 1200 KHz while the other was connected to a narrow bandpass preamplifier (400 - 600 KHz). Two different ranges of bandpass filter frequency for the amplifier in the LOCAN-AT instrumentation were used: 100 - 300 KHz and 400 - 600 KHz. Gain, threshold, peak definition time (PDT), hit definition time (HDT) and hit lockout time (HLT) values were set at 20 db, 45 db, 20 μ s, 150 μ s and 300 μ s respectively. These variables controlled the signal measurement process.

Application of AE proved rather adept to coupon testing as the transducers were easily mounted to the specimens using masking tape and rubber bands. A small amount of vacuum grease was placed between the transducers and the specimen to ensure air gaps would not permit acoustic wave reflection interference. Both transducers occupied one side of the specimen and were spaced about 1.5" apart center to center. In general, AE monitoring for a given specimen was initiated before startup of the mechanical testing instrument so that all AE would be included for analysis.

Finally, temperature build-up as a function of fatigue lifetime was recorded with a thermocouple that was attached to the specimen with a binder clip. To minimize the possibility of external noise caused by friction between clip and specimen surface, a piece of rubber was placed in between the contact areas. The location of the thermocouple was generally between the transducers but on the opposing side of the specimen.

RESULTS

The results of the experimentation with tire coupons appear as key points and relate to Figures 2 through 8e. as well as those in Appendix A.

Fatigue lifetime of the composite model system is controlled by stress amplitude rather than maximum stress as shown by the comparison of Figures 2 and 3. Figure 3 particularly highlights this as the curves are shown to converge with the exception of the case where minimum stress approaches zero. This series of plots also illustrates that at a given stress amplitude, the effect of R-ratio (minimum load/maximum load) on fatigue lifetime is negligible. Again, an exception is evident as the minimum stress approaches zero.

The fact that shorter fatigue lifetime is observed when the minimum stress is near zero may be attributed to the alteration of failure modes. Normally, cord-rubber composites exhibit debonding of cord from the surrounding rubber matrix developing into delamination. When the minimum stress approaches zero, microbuckling of laminates may occur; as a result, the delamination process could precede cord-rubber debonding. However there was no concrete evidence of this based on visual observation during testing.

The correlation between dynamic creep rate and fatigue lifetime is given in Figure 4. The plot indicates that fatigue lifetime is inversely proportional to the rate of dynamic creep according to a power law.

Temperature effects on fatigue lifetime are shown in Figures 5 and 6. Figure 5 establishes the relation between rate of temperature increase and fatigue lifetime. Again, fatigue lifetime is shown to be inversely proportional to temperature rate increase. However, as Figure 6 firmly suggests, there is

no correlation between fatigue lifetime and final temperature of the specimen just prior to catastrophic failure.

Under static loading, strain dependence of AE events, counts and energy which reflects changes of failure mode has been clearly established. Figure 7 illustrates this correlation using the dependent variable of acoustic energy. The plot is a compilation of four different specimens where a definite trend is obvious supported by the good agreement between specimens. As shown, there is a distinct change in slope corresponding directly to changes in failure mode from damage accumulation. The portion of the curve with high slope corresponds to delamination failure mode while that with smaller slope indicates debonding processes. Similar plots of events and counts versus axial strain show similar trends, but they have not been included in this report. Typical output of raw AE from static loading is presented in Figures 8a - 8e.

Unlike the case of static loading, dynamic loading produced less obvious trends in strain dependence of AE. Distinctly different rates of AE signal accumulation can be assigned to the delamination and cord-rubber debonding failure modes only when the minimum cyclic stress is high enough. The reason for this is unclear at the present time. Plots of select specimens that did exhibit trends in AE as a function of strain were not readily available at the time of this write-up, however preliminary analysis of the raw AE output in Appendix A (Figures 1 through 3) indicates trends are apparent, selective as they may be.

CONCLUSION

Using the results acquired from Summer 1990 as a baseline, the work

completed this year has provided the necessary information that is critical for a growing and reliable data base. Results show that damage accumulation and fatigue lifetime is more a function of stress amplitude rather than mean stress at a given cycling frequency of 5 Hz.

Although the relationship between dynamic loading and AE behavior was less definitive than that for static loading, the potential of using AE as a real-time monitoring technique for aircraft tires still has some merit. Enhanced signal processing and analysis seems to be the key in reproducing useful acoustic emission that has no interference from external noise sources. For example, gating circuitry could eliminate gripping noise (from the mechanical fixtures that hold specimens in place during testing) by selectively and instantaneously measuring acoustic signals when no load is applied to the fatigued specimen. Also experimenting with instrument settings may offer some refinements that were not investigated with this study. This would apply mostly to the threshold amplitude setting. Increasing this value to say 70 db may prove helpful.

Table 1. Cyclic Loading Combinations for Cord-Rubber Model
Composites

FileName	Specm#	Width	Pmax	Pmin	Freq
BWAG1102	493-11-2	.712"	800lbf	100lbf	5Hz
BWAG1201	493-12-1	.747	800	100	5
BWAG1301	493-13-1	.775	800	200	5
BWAG1401	493-14-1	.722	800	200	5
BWAG1501	493-15-1	.702	800	400	5
BWAG1601	493-16-1	.737	800	400	5
BWAG1701	493-17-1	.719	800	0	5
BWAG1801	493-18-1	.721	800	0	5
BWAG1901	493-19-1	.714	700	0	5
BWAG2001	493-20-1	.733	700	0	5
BWAG1202	493-12-2	.733	700	100	5
BWAG1302	493-13-2	.727	700	100	5
BWAG1402	493-14-2	.738	700	200	5
BWAG1502	493-15-2	.742	700	200	5
BWAG1602	493-16-2	.746	600	0	5
BWAG1702	493-17-2	.732	600	0	5
BWAG1802	493-18-2	.711	600	100	5
BWAG1902	493-19-2	.727	600	100	5
BWAG2002	493-20-2	.722	600	200	5
BWAG0803	493-8-3	.741	600	200	5
BWAG1003	493-10-3	.743	710	390	5
BWAG1103	493-11-3	.740	710	390	5
BWAG1203	493-12-3	.763	830	570	5
BWAG1303	493-13-3	.746	830	570	5
BWAG1403	493-14-3	.729	500	0	5
BWAG1503	493-15-3	.723	500	0	5
BWAG1603	493-16-3	.710	500	100	5
BWAG1803	493-18-3	.770	500	100	5
BWAG1903	493-19-3	.730	400	0	5
BWAG2003	493-20-3	.739	400	0	5
BWAG0801	493-8-1	.733	750	250	5
BWAG1001	493-10-1	.794	750	250	5

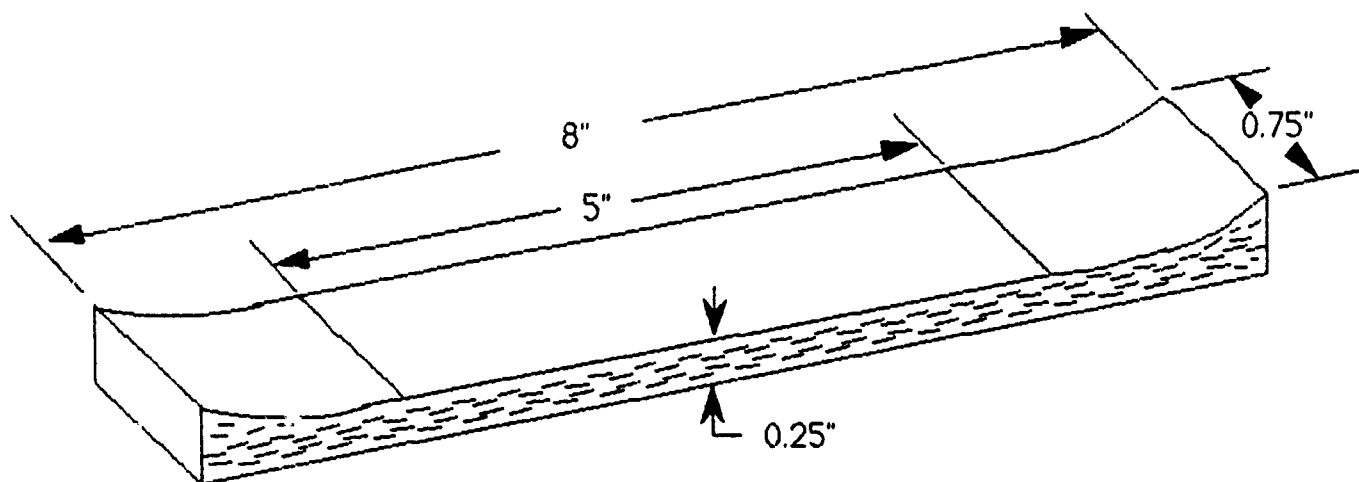


FIGURE 1. COUPON SCHEMATIC

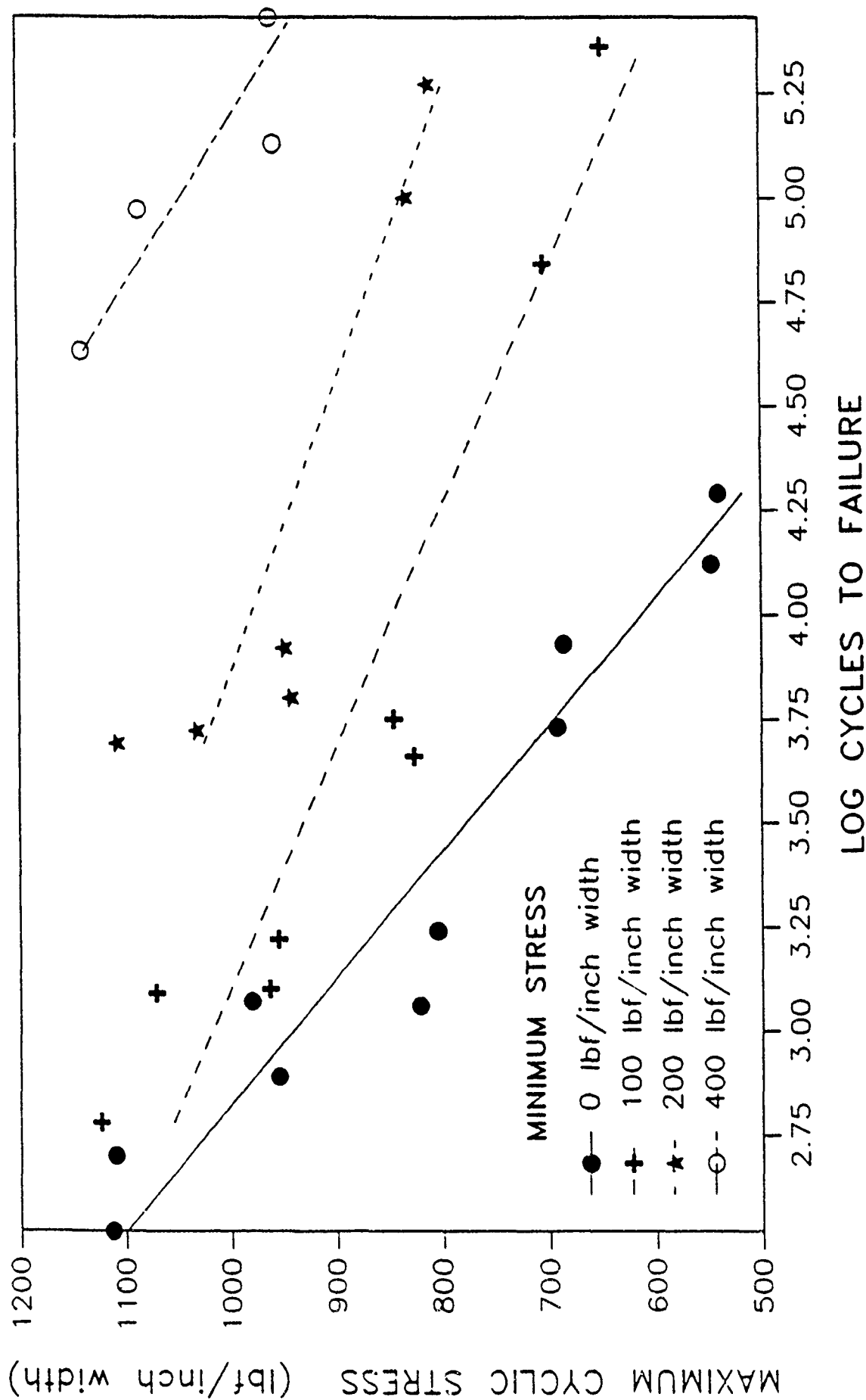


FIGURE 2. S-N CURVE FOR CORD-RUBBER COMPOSITES

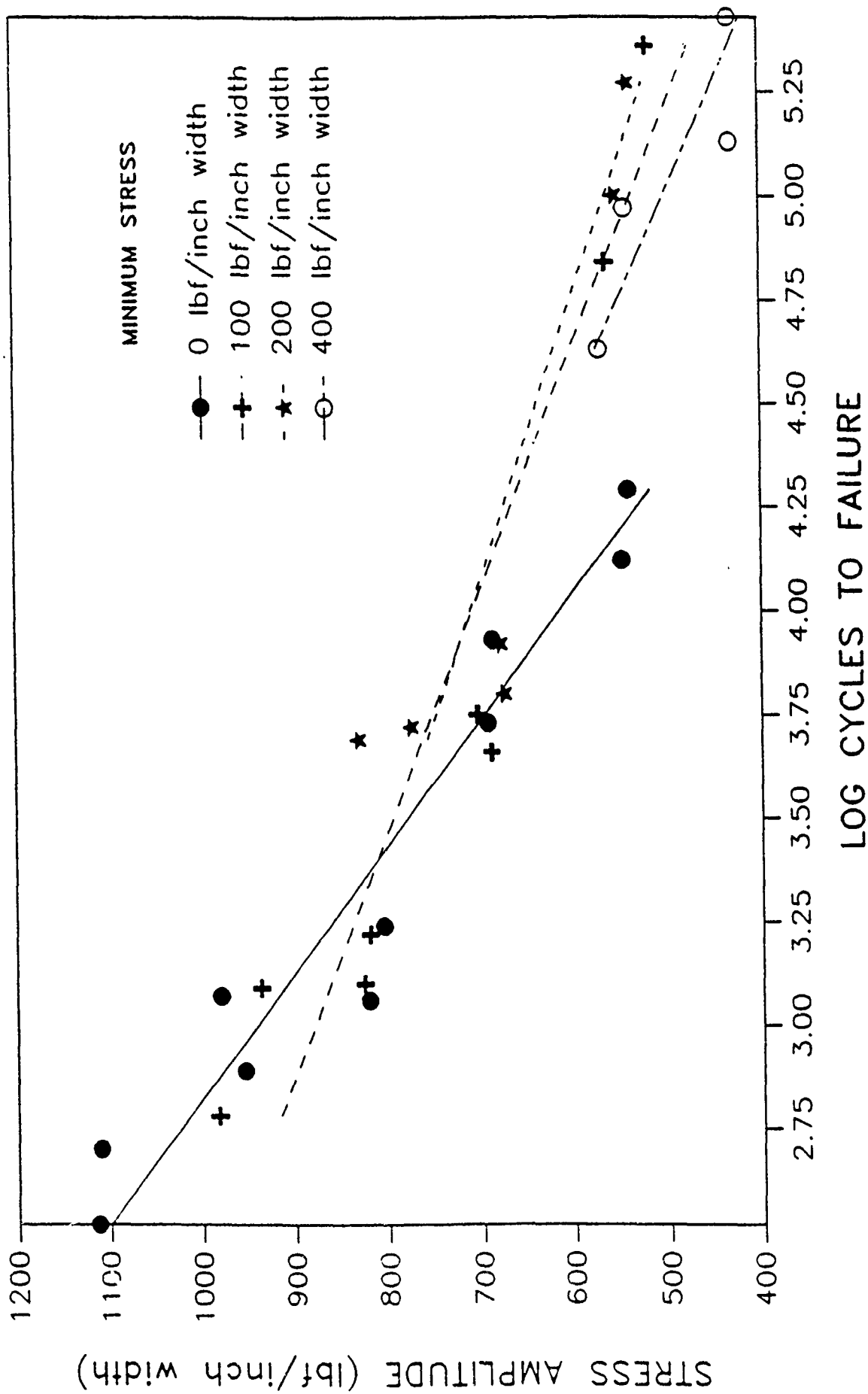


FIGURE 3. S-N CURVE FOR CORD-RUBBER COMPOSITES

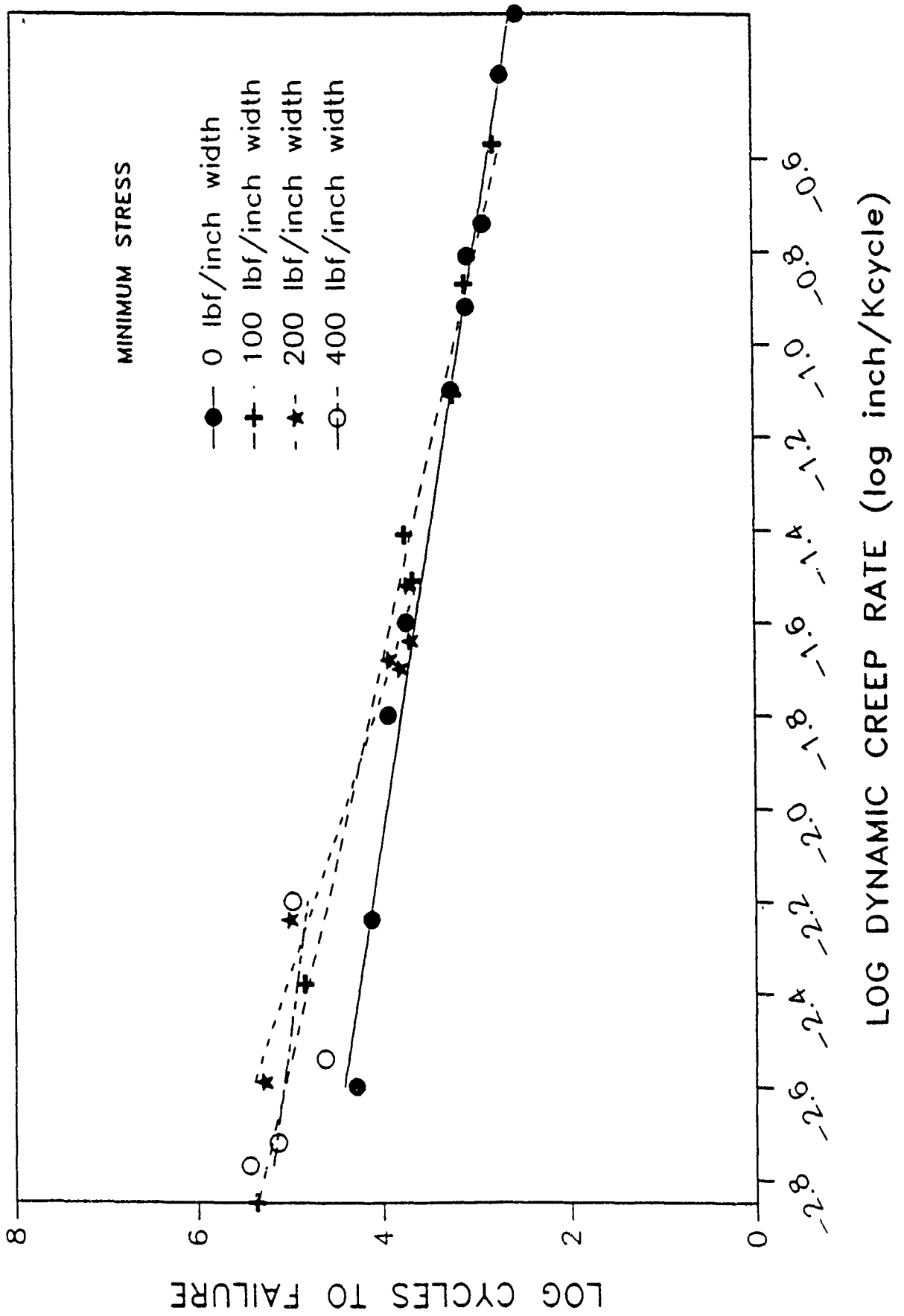


FIGURE 4. FATIGUE LIFETIME vs DYNAMIC CREEP RATE

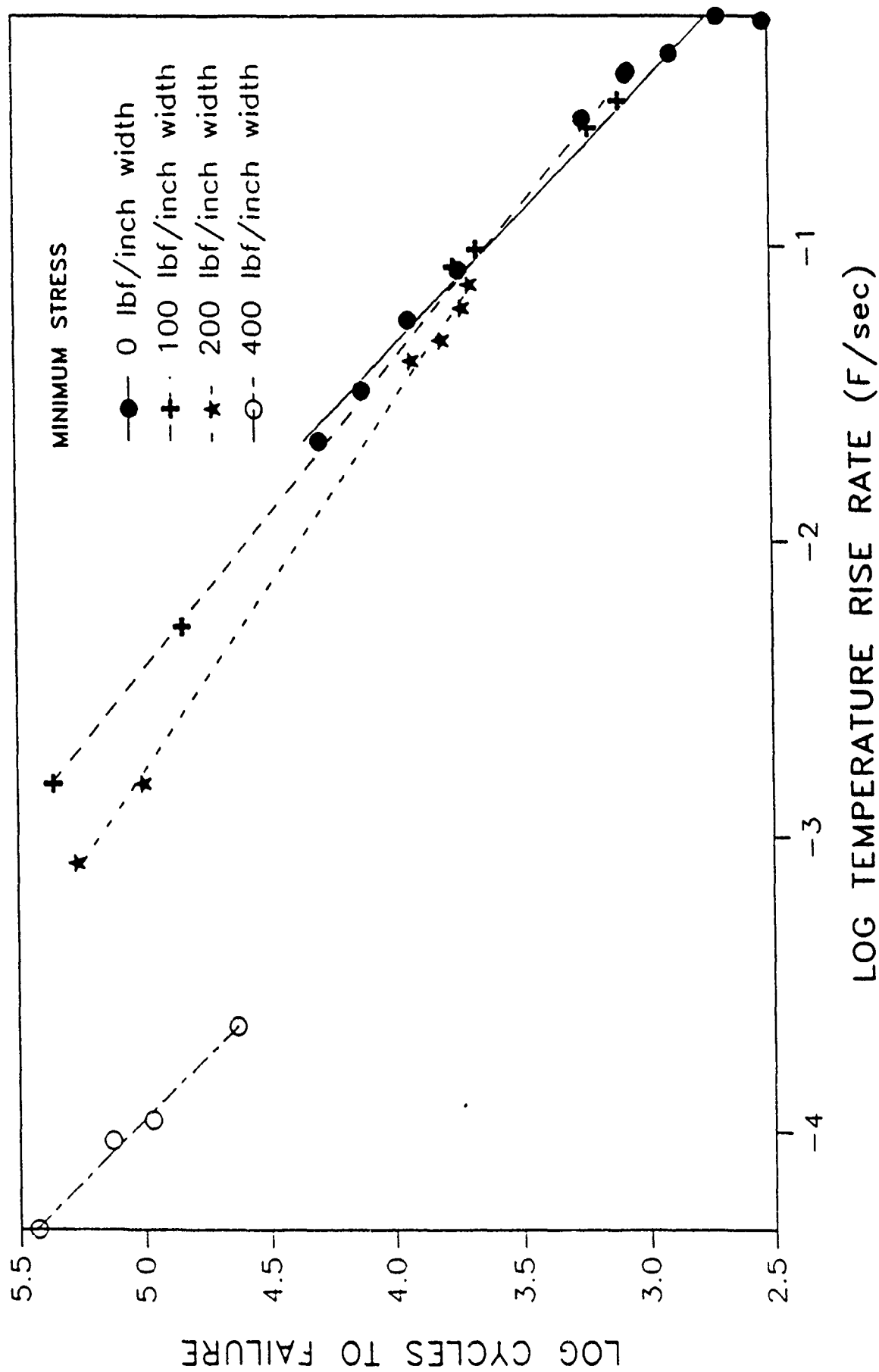


FIGURE 5. FATIGUE LIFETIME vs TEMPERATURE RISE RATE

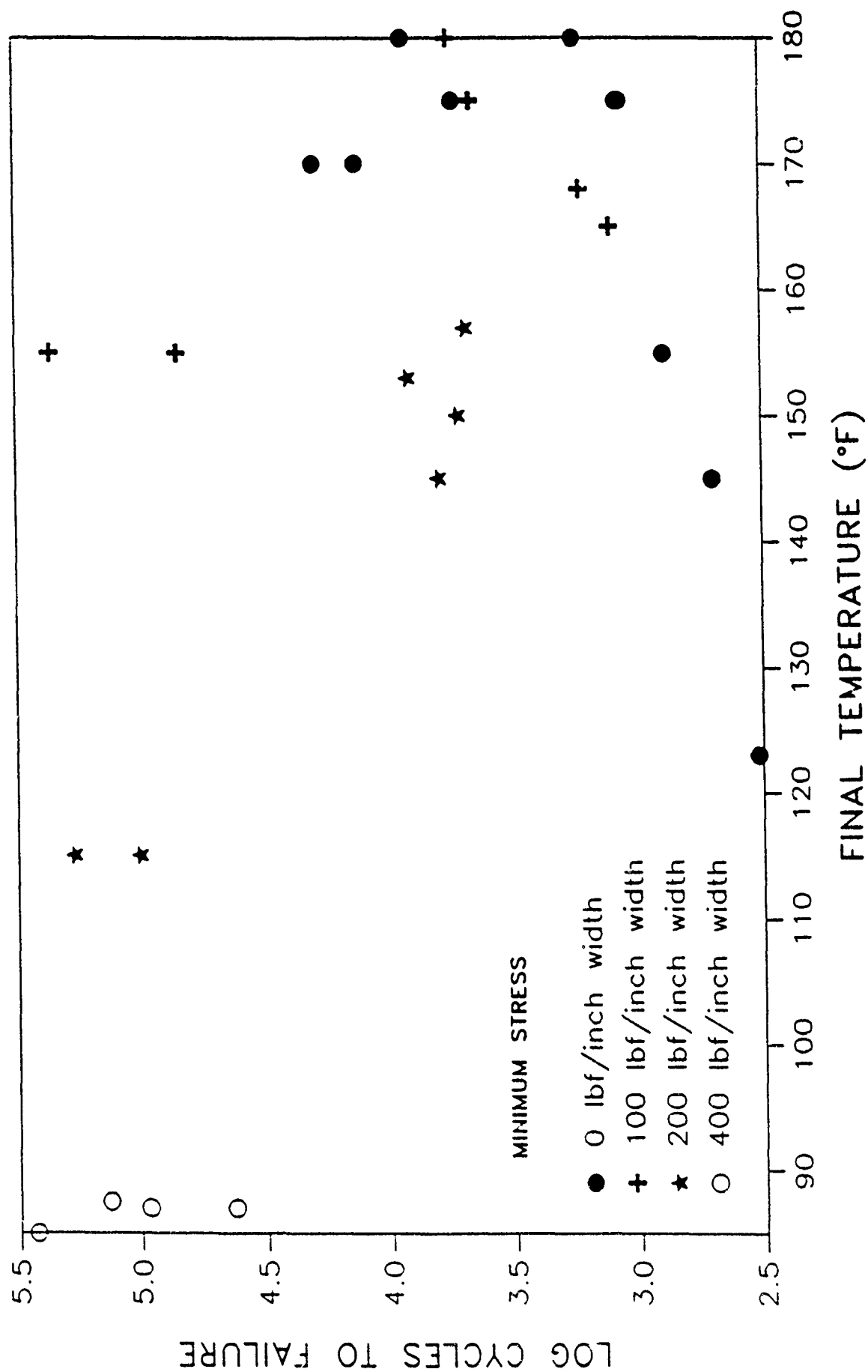


FIGURE 6. FATIGUE LIFETIME vs FINAL TEMPERATURE

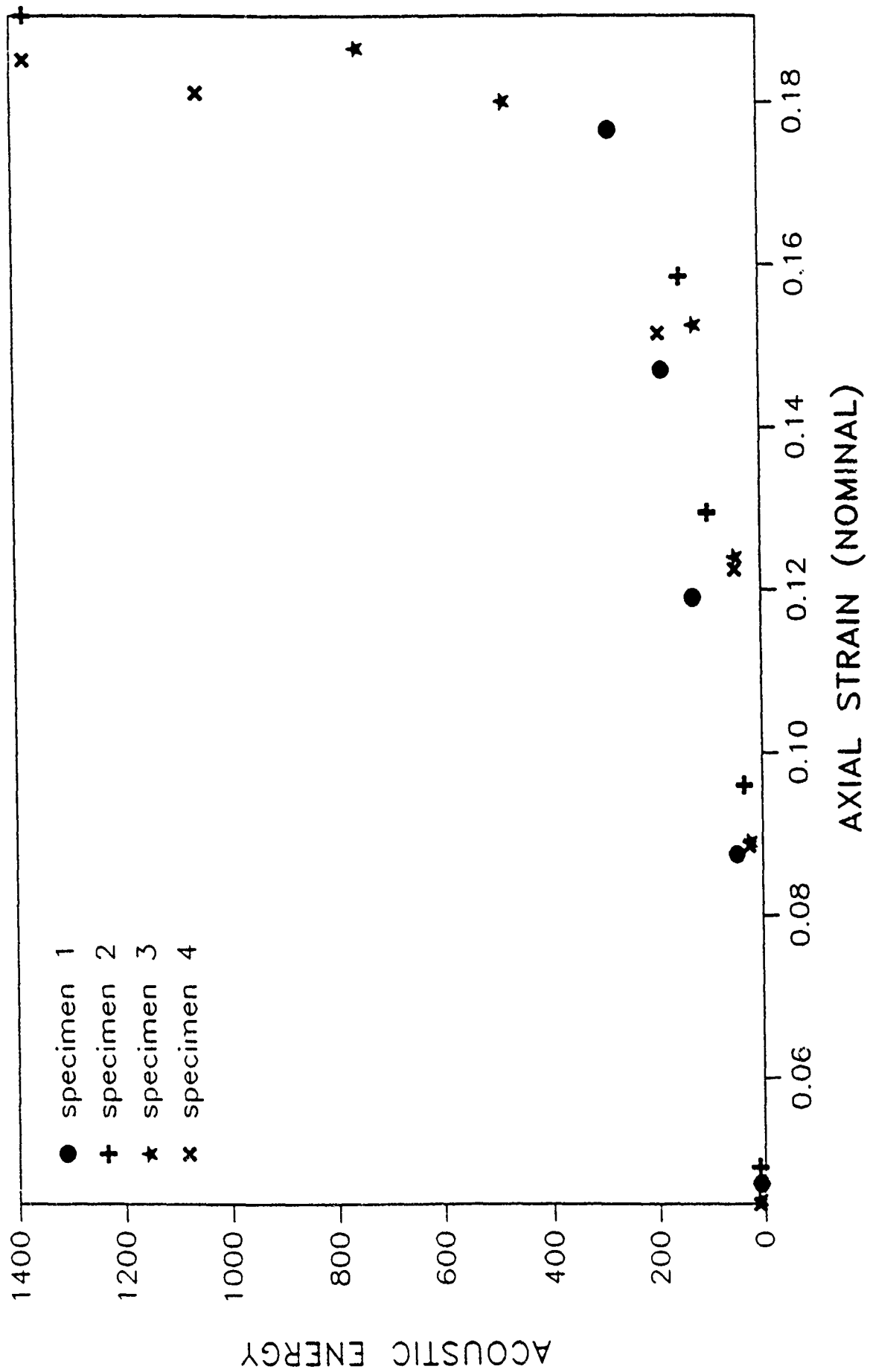
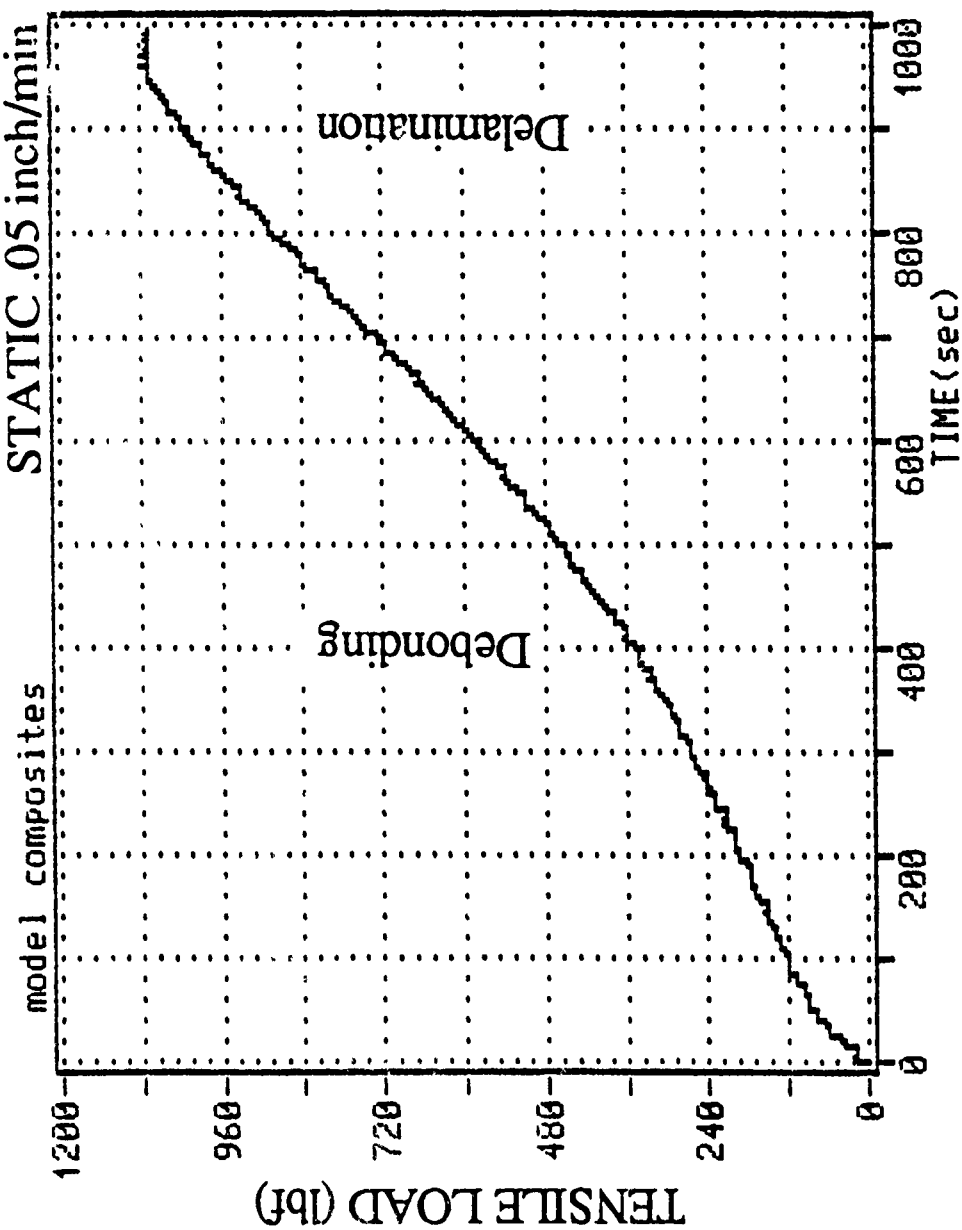


FIGURE 7. ACOUSTIC ENERGY vs AXIAL STRAIN (STATIC LOADING)



AE HITS	LOST AE
524	
CUM-ENR	CUM-ENR
6764	
0000 HH:MM:SS	
0 00:18:43	
LOAD #1	CYCLE-C
5.00	
1E4	
1E2	
1	
HITS VS CHANNEL	
F5 PRINT GRAPH	
F6 USER COMMENT	
F7 PREV. GRAPH	
F8 NEXT GRAPH	
F9 STOP	
F10 TO CANCEL	
# <CR> = GRAPH	
REPLAY END	
A:ACORDEEEET.DTA	

FIGURE 8a. PARAMETRIC OUTPUT (LOAD VS TIME)

2AE HITS	LOST AE
524	
CUM-ENR	
6764	
000:HH:MM:SS	
0 00:18:43	
LOAD #1	CYCLE-C
5.00	

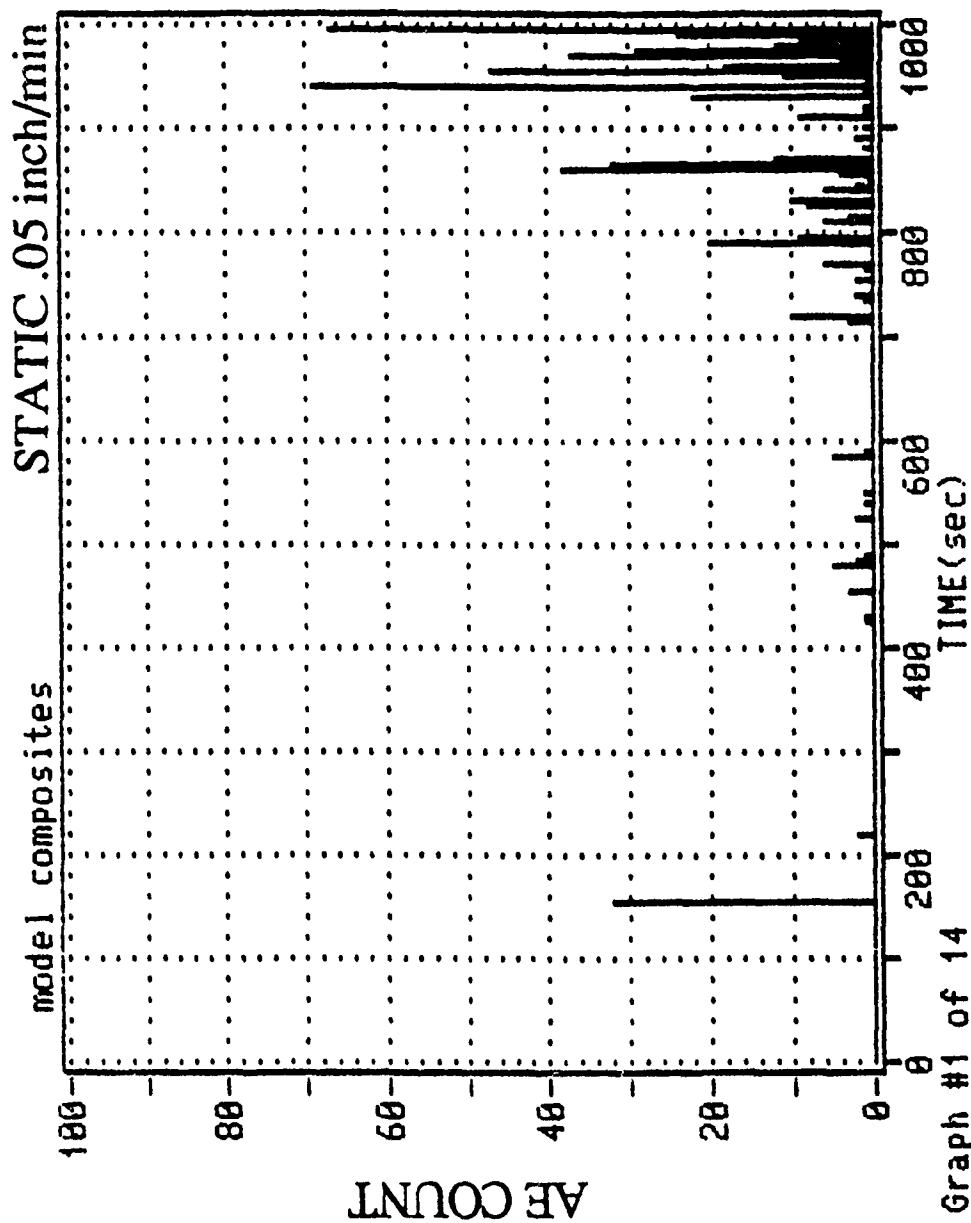
1E4	
1E2	
1	

HITS vs CHANNEL

F5 PRINT GRAPH
 F6 USER COMMENT
 F7 PREV. GRAPH
 F8 NEXT GRAPH
 F9 STOP

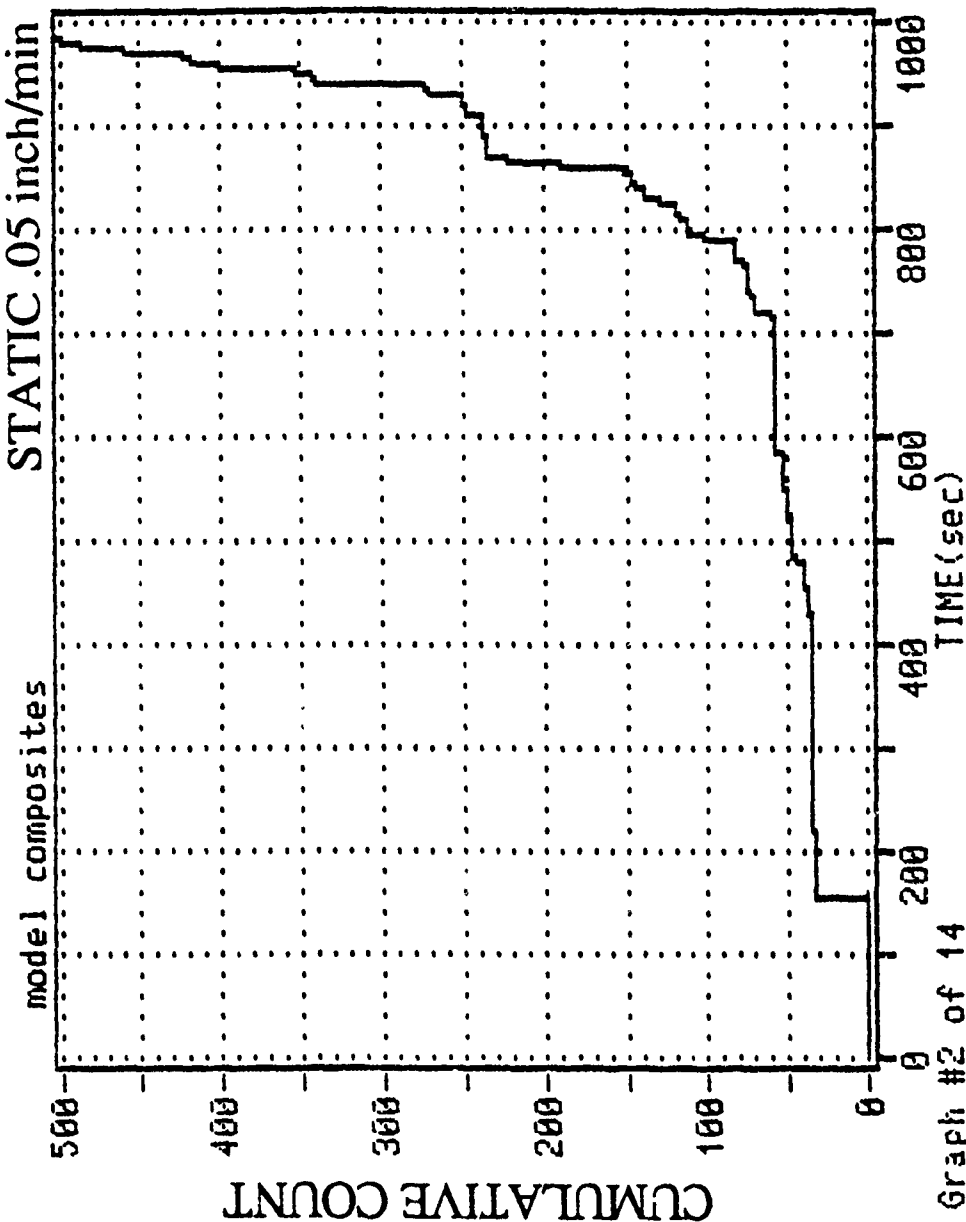
F10 TO CHANNEL

<CR> = GRAPH
 REPLAY END
 A:CORD0001.DTA



400-600 KHz Bandpass Filter

FIGURE 8b. AE COUNT vs TIME



```

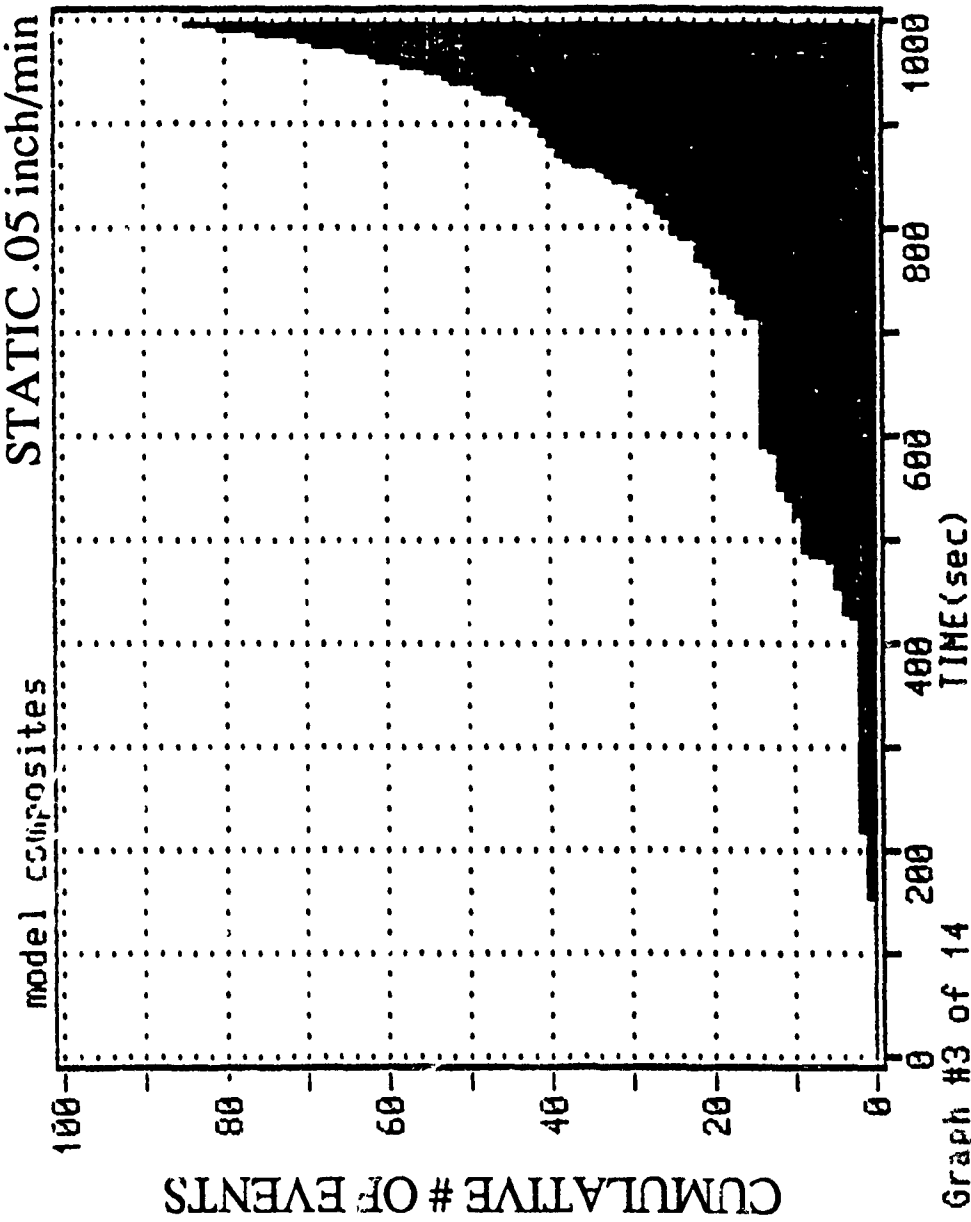
NAE-HITS LOST AE
524
GUN=CHTS CUN-ENER
6764
ADD HHMMSS
0 00:18:43
LOAD HI CYCLE-C
5.00

1E4
1E2
1

HITS vs CHANNEL

F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
RECORD0000.DTA
  
```

FIGURE 8c. AE CUMULATIVE COUNT vs TIME



AE HITS	LOST AE
524	
CUM-HITS	CUM-EHER
6764	
DDD HH:MM:SS	
0 00:18:43	
WORD #1	CYCLE-C
5.00	

1E4-	
1E2-	
1-	

HITS vs CHANNEL

F5 PRINT GRAPH

F6 USER COMMENT

F7 PREV. GRAPH

F8 NEXT GRAPH

F9 STOP

F10 TO CANCEL

<CR> = GRAPH

REPLAY END

AE-CORRECT 014

FIGURE 8d. AE CUMULATIVE EVENTS vs TIME

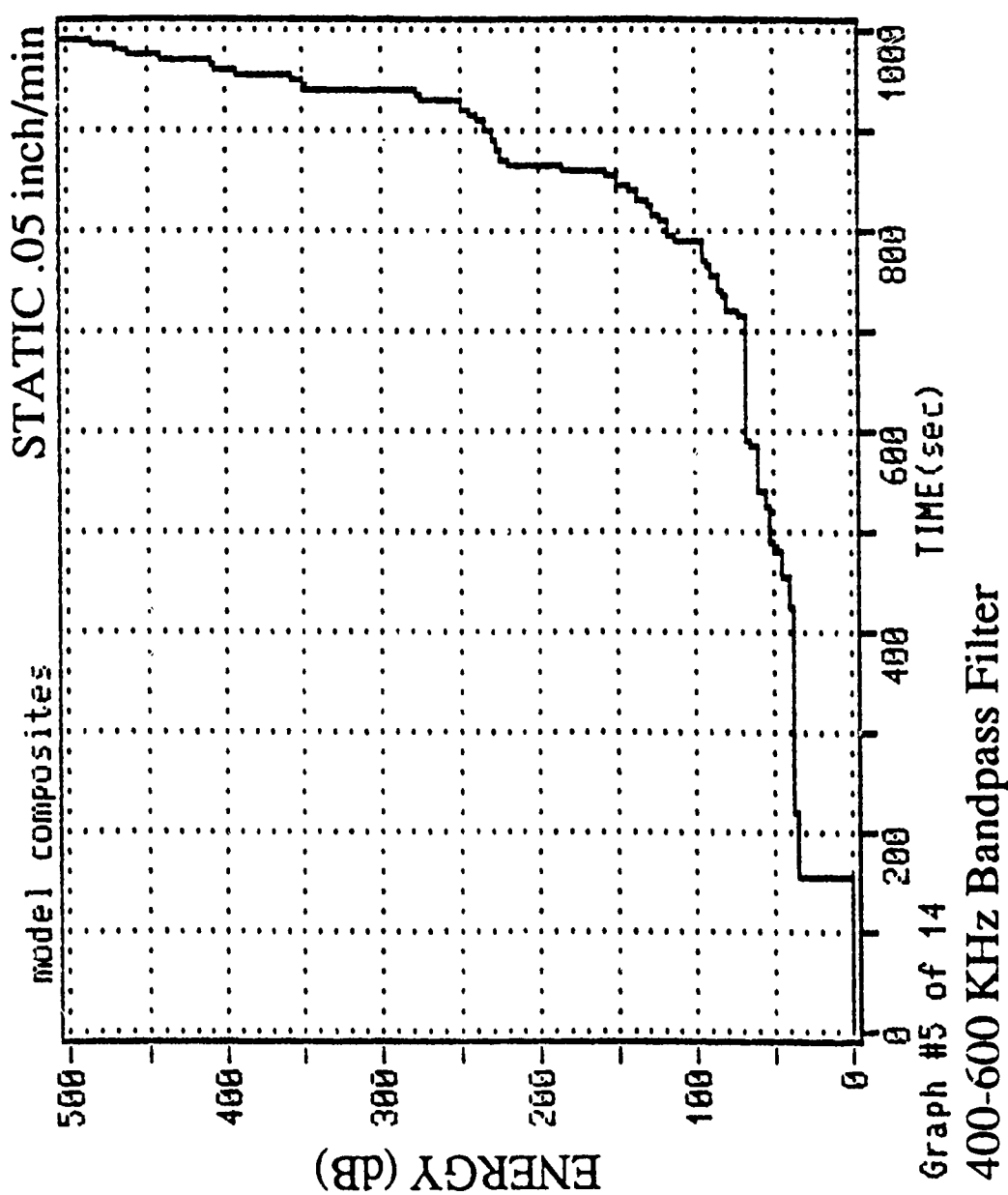


FIGURE 8e. AE CUMULATIVE ENERGY vs TIME

AE HITS	LOST AE
524	
CUR-CNTS	CUM-ENER
6764	
0000 HH:MM:SS	
0 00:18:43	
LOAD HI	CYCLE-C
5.00	
1E4	
1E2	
1	
HITS vs CHANNEL	
F5 PRINT GRAPH	
F6 USER COMMENT	
F7 PREV. GRAPH	
F8 NEXT GRAPH	
F9 STOP	
F10 TO CANCEL	
# <CR> = GRAPH	
REPLAY END	
AE00000001.DTA	

APPENDIX A: RAW AE DATA FROM CYCLIC LOADING OF CORD-RUBBER
MODEL COMPOSITES

```

AE-HITS LOST HE
744
CUM-ENTS CUM-ENER
5553
000 HH:MM:SS
9 32:12:16
LOAD-#1 CYCLE-1
5.50
154-
122-
1-
HITS 16 CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A-BRAC16B1-01A-2

```

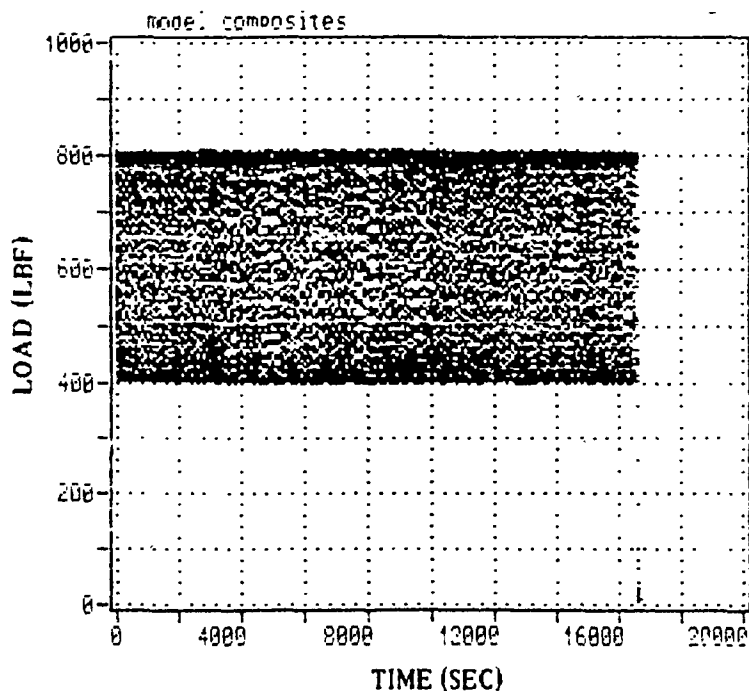


FIGURE A1a. CYCLIC LOAD vs TIME

```

AE HITS LOST RE
741
CUM-ENER
5552
000 HH:MM:SS
0 22:12:16
ALOAD #1 CYCLE-C
5.50

1E4
1E2
1
HITS vs CHANNEL

F5 PRINT GRAPH
F6 USER COMMENT
F7 PPREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A:BNAG1507.DTA

```

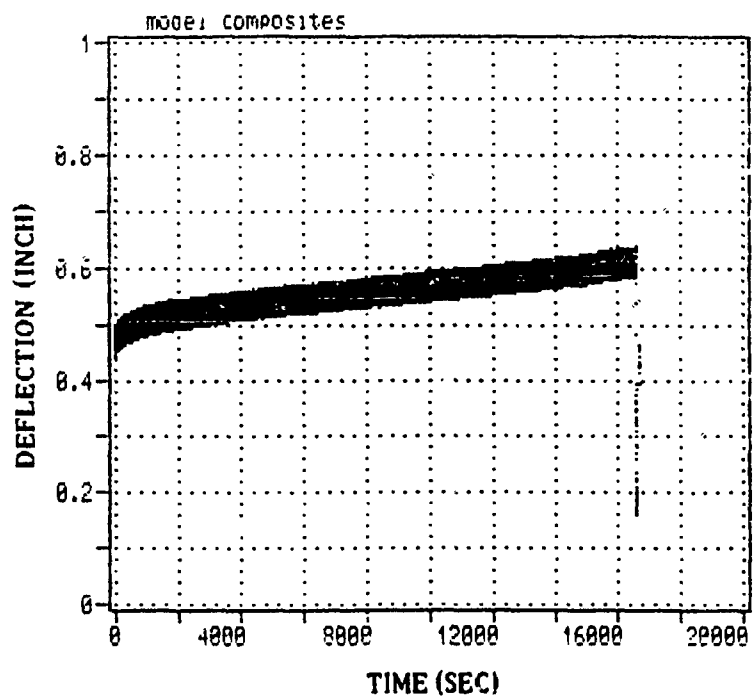


FIGURE A1b. DEFLECTION vs TIME

```

AE HITS LOST AE
741
CUM-ENTS CUM-ENER
5553
EOD HH:MM:SS
0 22:12:16
COPD #1 CYCLE-C
5.50
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
#00001600, 0TA

```

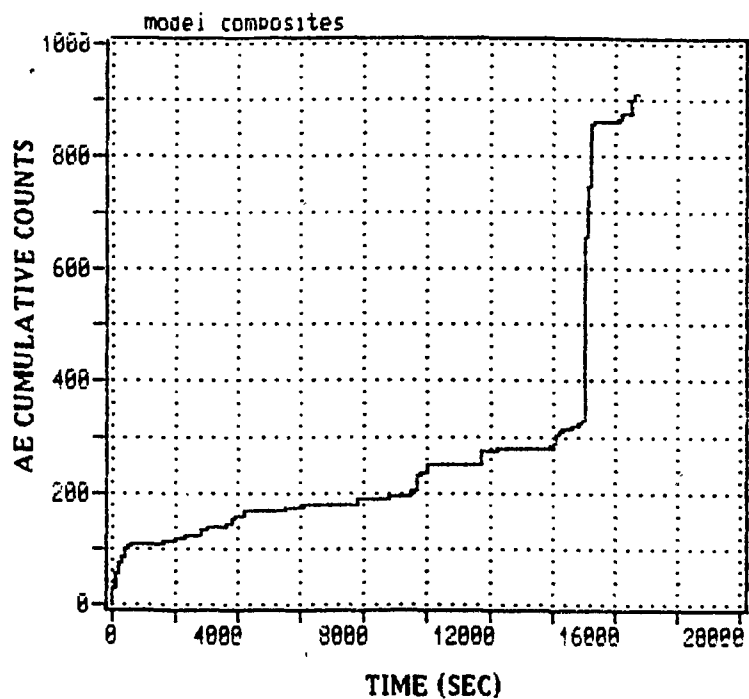


FIGURE A1c. AE COUNTS vs TIME

```

AE HITS LOST AE
741
CUM-CRCS CUM-ENER
5553
DDD HH:MM:SS
0 22:12:16
LOAD #4 CYCLE-C
5.50
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# (CR) = GRAPH
REPLAY END
AEHW-6150*.DTA

```

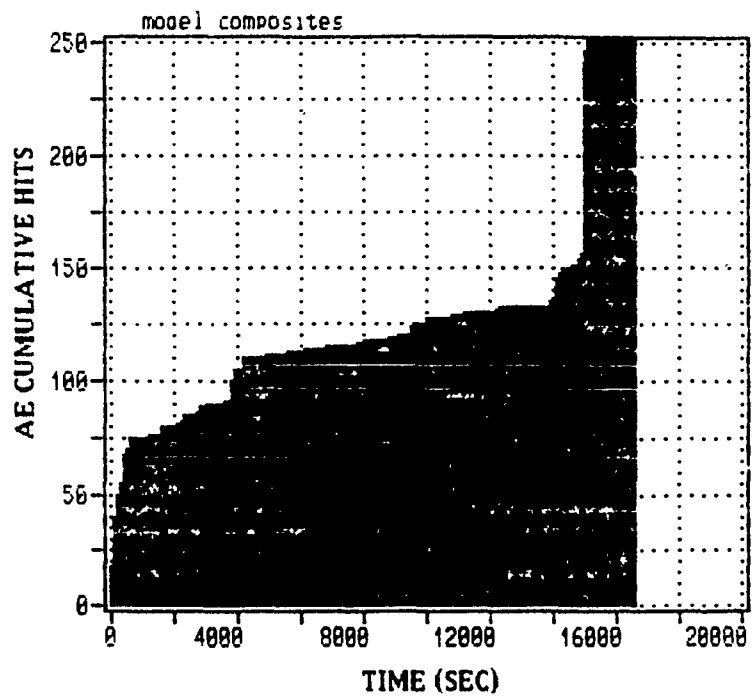


FIGURE A1d. AE HITS vs TIME

```

AE HITS LOST AE
741
CUM-ENTS CUM-ENER
5553
000 22:12:33
0 22:12:16
LOAD #1 CYCLE-C
5.58
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
# <CR> = GRAPH
REPLAY END

```

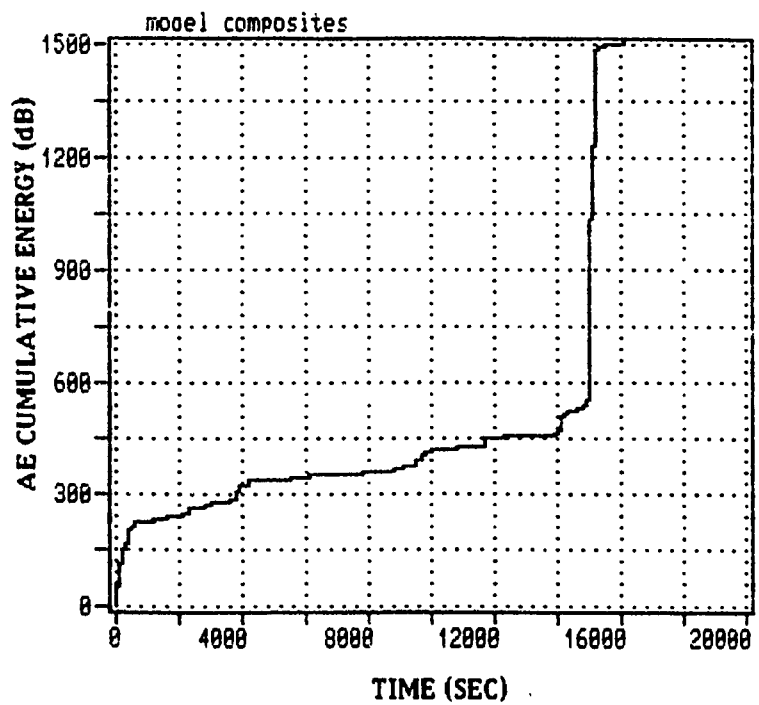


FIGURE A1e. AE ENERGY vs TIME

```

HE HITS LOST AE
741
CUM-HITS CUM-ENER
5553
000 HH:MM:SS
0 22:12:16
LOAD #1 CYCLE-C
5.50
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CHANNEL
# <CR> = GRAPH
REPLAY END
APR04-1984, DT-

```

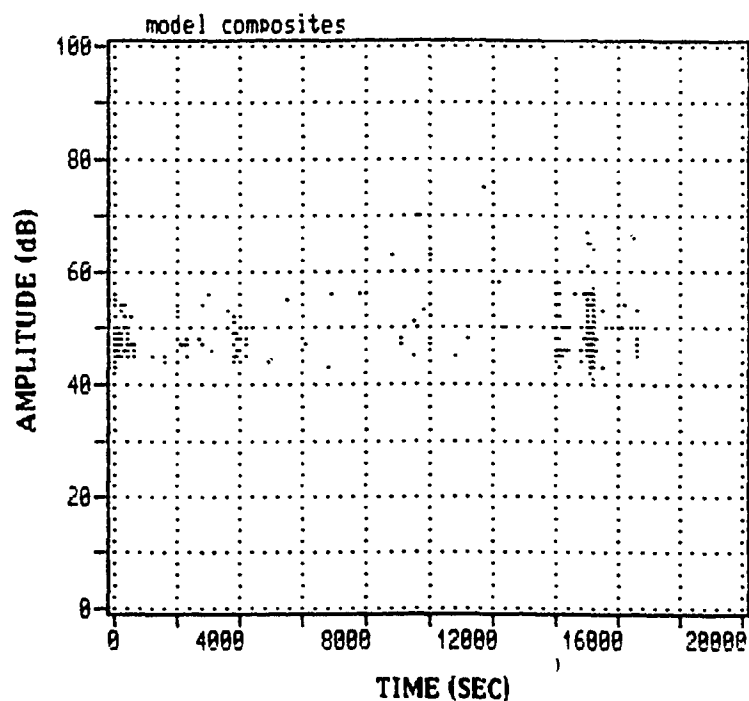


FIGURE A1f. AMPLITUDE vs TIME

```

PRE-HITS   LOST RE
366
CUM-ONTS   CUM-ENER
38551
0000 HH:MM:SS
0 03:24:41
LOAD #1    CYCLE-C
8.00

1E4
1E2
1
HITS vs CHANNEL

F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A:6048-501V.DTA

```

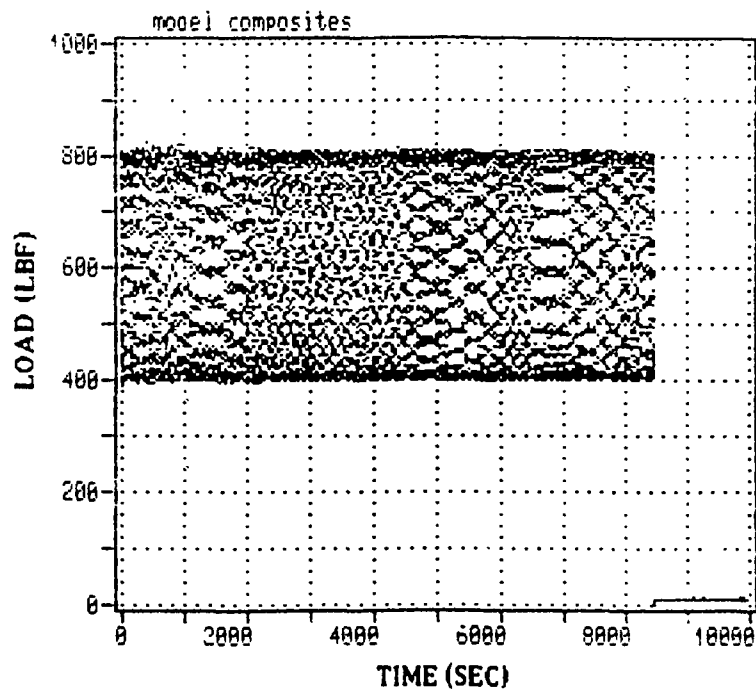


FIGURE A2a. CYCLIC LOAD vs TIME


```

AE HITS LVST ME
366
CUM-CNTS CUM-ENER
38551
000 HH-MM-SS
0 03:24:41
WPR0-#1 CYCLE-0
8.00
154
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PPEV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
H:BNAB1501.0TP

```

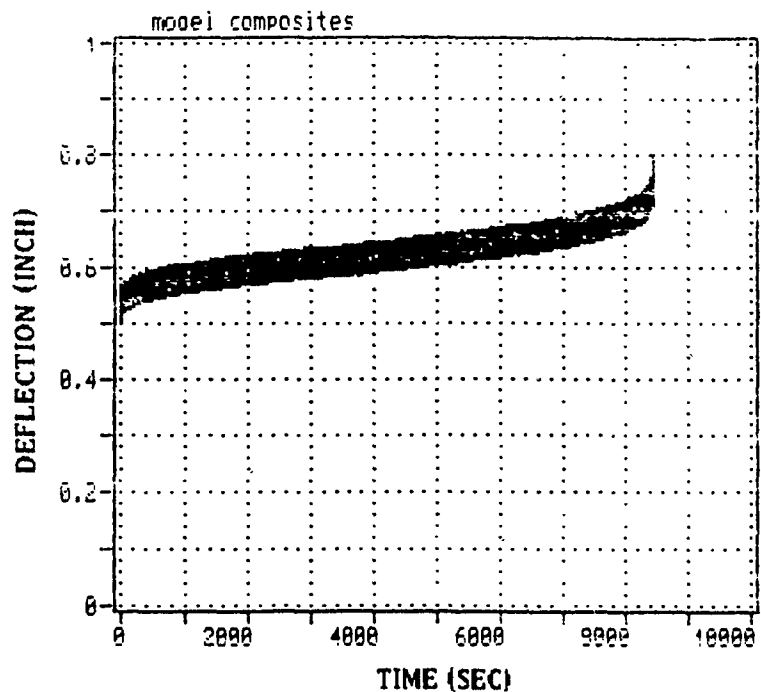


FIGURE A2b. DEFLECTION vs TIME

```

AE HITS  LVST MEI
366
CUM-CNTS CUM-ENER
38551
000:HH:MM:SS
0 03:24:41
LOAD #1 CYCLE-C
8.00
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A:BNAC1501.OTA

```

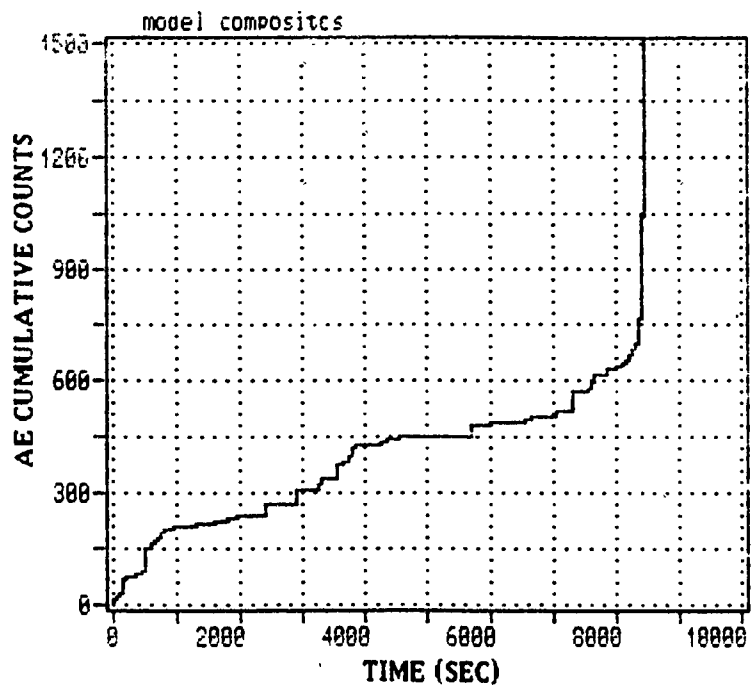


FIGURE A2c. AE COUNTS vs TIME

```

AE HITS LOST AE
366
CUM-HITS CUM-ENER
38551
00:00:00 HH:MM:SS
0 03:24:41
LOAD-SET CYCLE-C
8.00
1E4
1E2
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PPEV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# (CR) = GRAPH
REPLAY END
A=SWAG1501.DTA

```

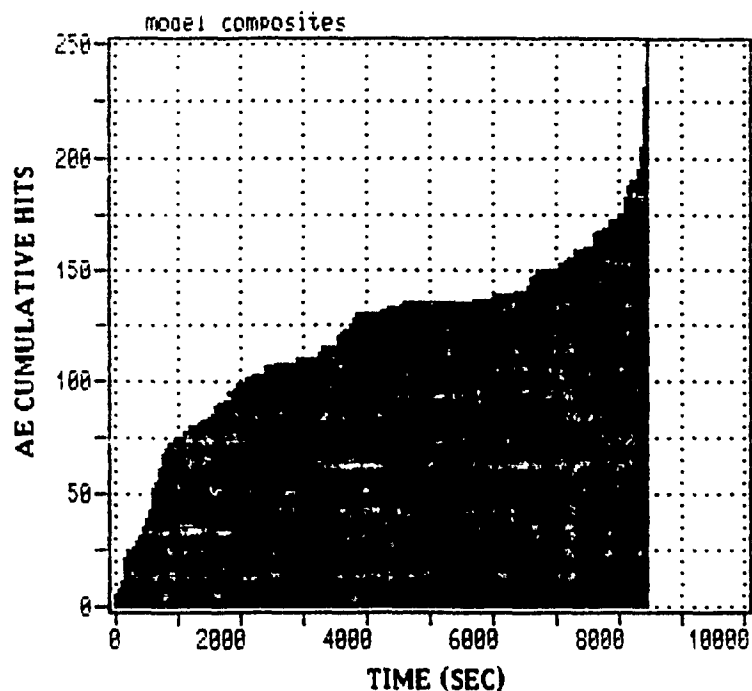


FIGURE A2d. AE HITS vs TIME

```

AE-HITS  LVST HEI
366
CUM-HITS CUM-ENER
38551
000 HH:MM:SS
0 03:24:41
LVST #1 CYCLE-C
0.00
1E4-
1E2-
1-
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TOX-CANCEL
# <CR> = GRAPH
REPLAY END
AE-BWAG1501.01A

```

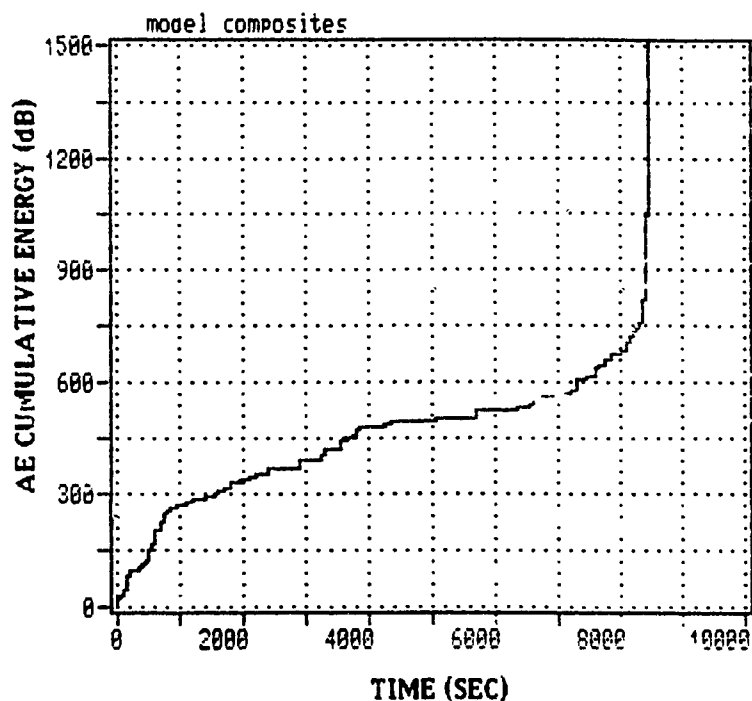


FIGURE A2e. AE ENERGY vs TIME

```

RE HITS  LOST RE
366
CUM-ENTS CUM-ENER
38554
000 HH:MM:SS
0 03:24:41
LVR0-#1 CYCLE-C
8.00
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CHANNEL
# <CR> = GRAPH
REPLAY END
AUG1501:074

```

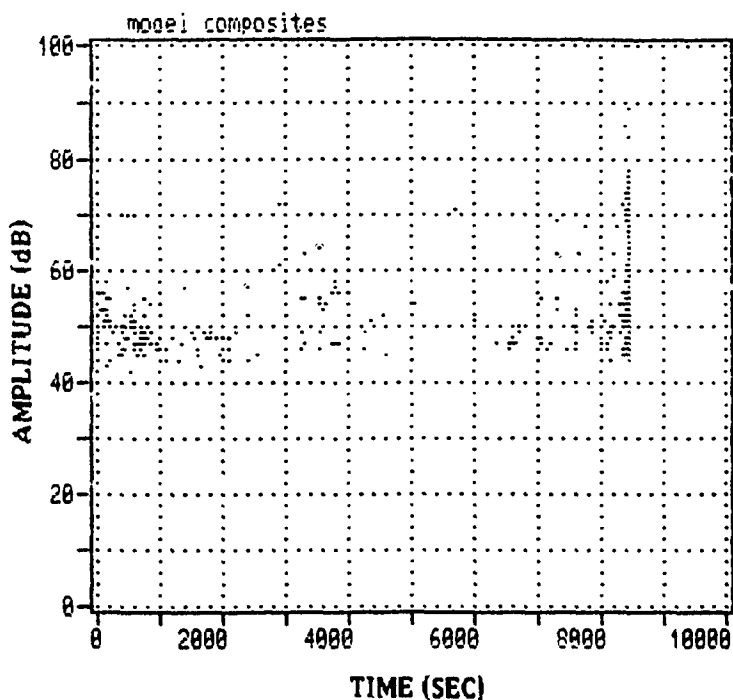


FIGURE A2f. AMPLITUDE vs TIME

```

RE HITS LOST HE
991
GUN-SENS CUM-ENER
4884
000 HH:MM:SS
0 01:16.39
LOAD #1 CYCLE-0
0.00
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
* <CP> = GRAPH
REPLAY END
A98W460001.0TA

```

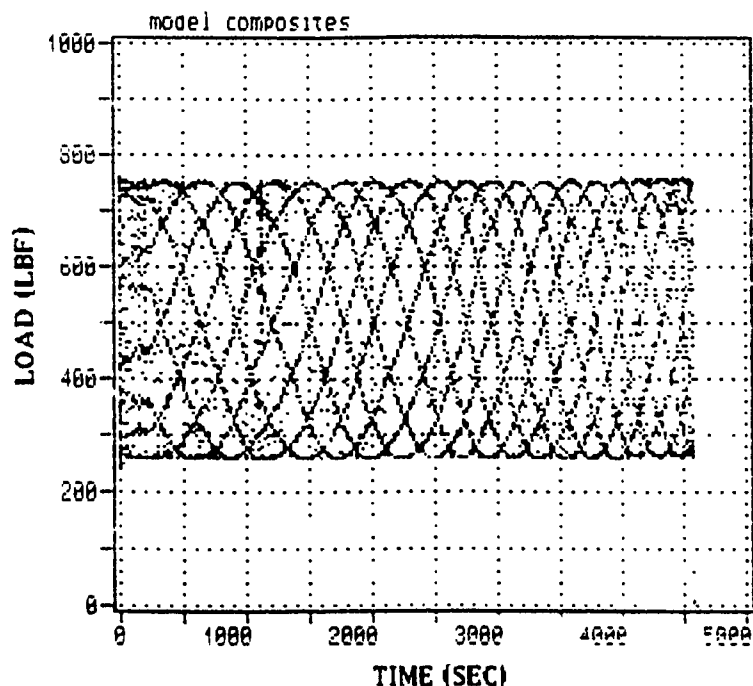


FIGURE A3a. CYCLIC LOAD vs TIME

```

ARE HITS LOST RE
994
CUM-HITS CUM-ENER
4884
000 HH:MM:SS
0 01:16:29
LOAD #1 CYCLE-C
8.00
1E4
1E2
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A-BMM9881.DTA

```

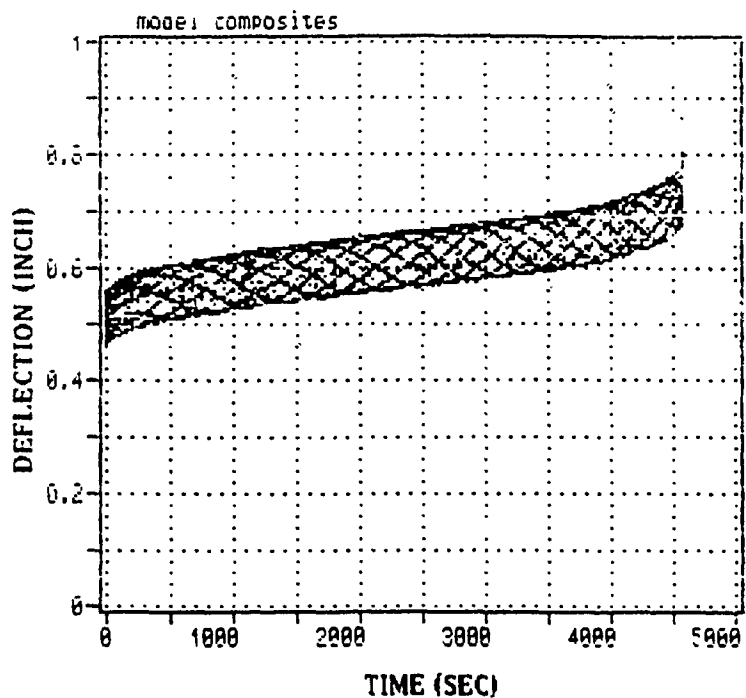
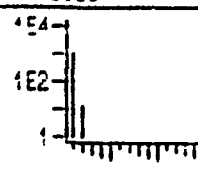


FIGURE A3b. DEFLECTION vs TIME

```

AE-HITS LOST ME
954
CUM-CNTS CUM-ENER
4884
000 HH-MM-SS
0 01:16:39
LEAD #1 CYCLE-0
8.08

```



```

HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# (CR) = GRAPH
REPLAY END
A:BN468804.DTA

```

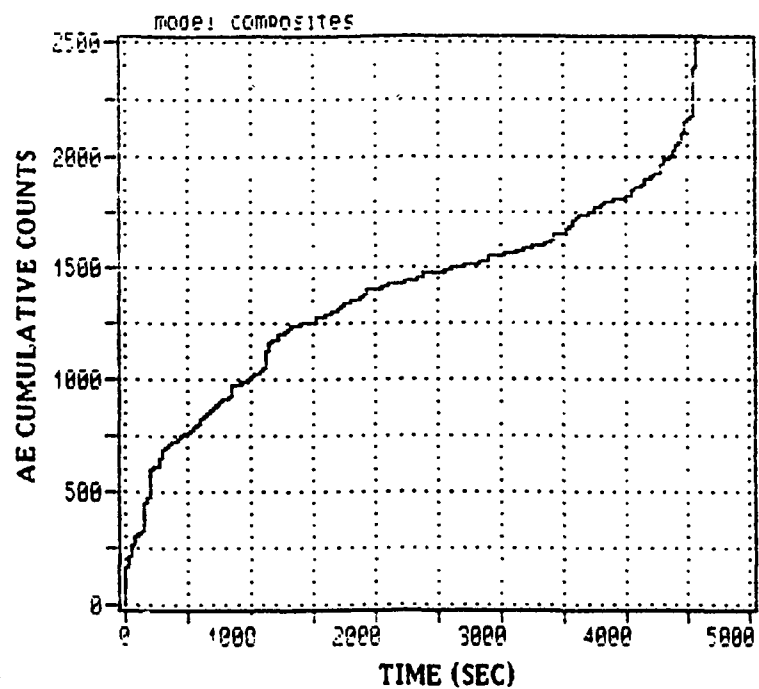


FIGURE A3c. AE COUNTS vs TIME


```

AE HITS LOST HE
991
CUM-CHITS CUM-ENER
4684
000 HH:MM:SS
0 01:16:39
LOAD-CHIT CYCLE-C
5.80
1E4
1E2
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
A-80460001.DTA

```

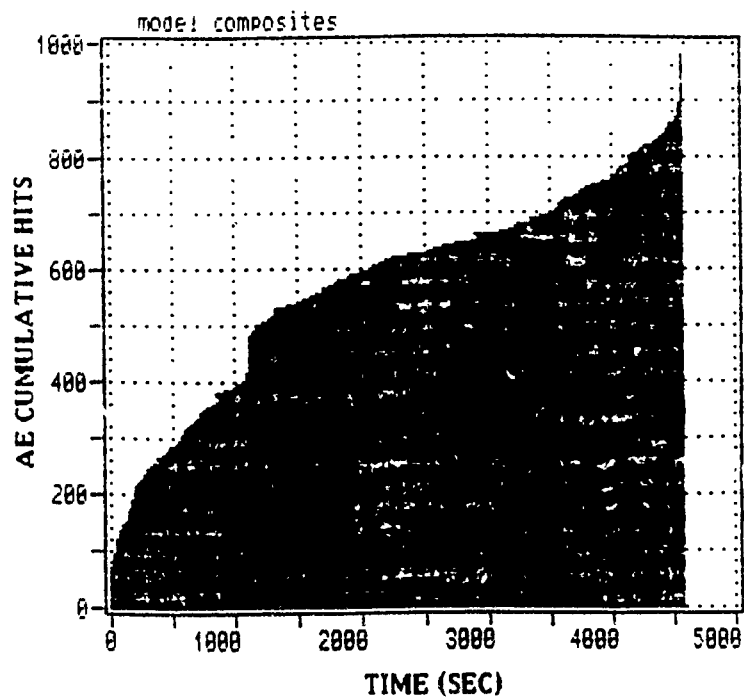


FIGURE A²⁴. AE HITS vs TIME

```

AE-STATS ... ME
394
CUM-CNTS CUM-ENER
4664
000 HH:MM:SS
8 01:16:39
LOAD #1 CYCLE-0
8.00
1E4
1E2
1
HITS vs CHANNEL
F5 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# <CR> = GRAPH
REPLAY END
(A: BWH-4545.DTA)

```

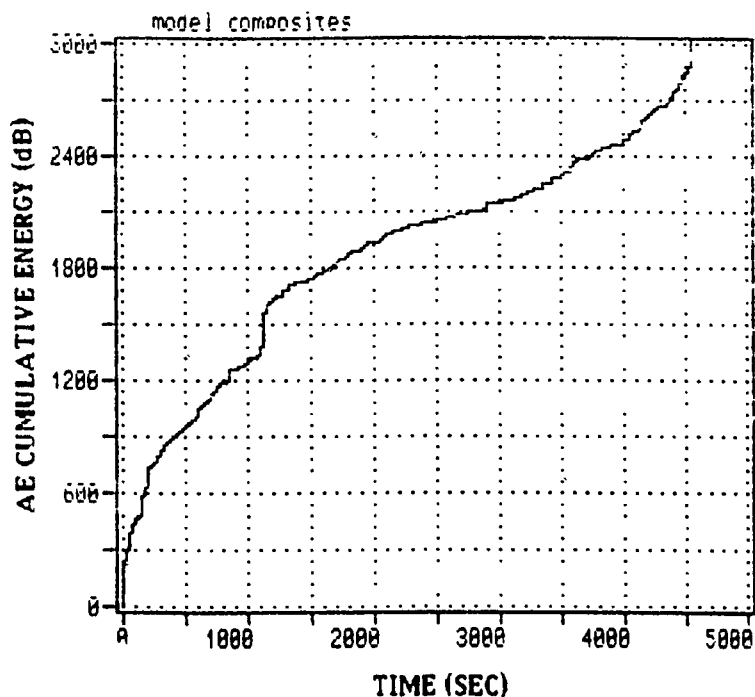


FIGURE A3e. AE ENERGY vs TIME

```

AF HITS  LOST RE
504
CUM-SNIS CUM-ENER
4664
000 HH:MM:SS
0 01:16:39
LONG-#1 CYCLE-C
3.00
154
1E2
1
HITS RE CHANNEL
F0 PRINT GRAPH
F6 USER COMMENT
F7 PREV. GRAPH
F8 NEXT GRAPH
F9 STOP
F10 TO CANCEL
# (CR) = GRAPH
REPLAY END
A:00000001.DT4

```

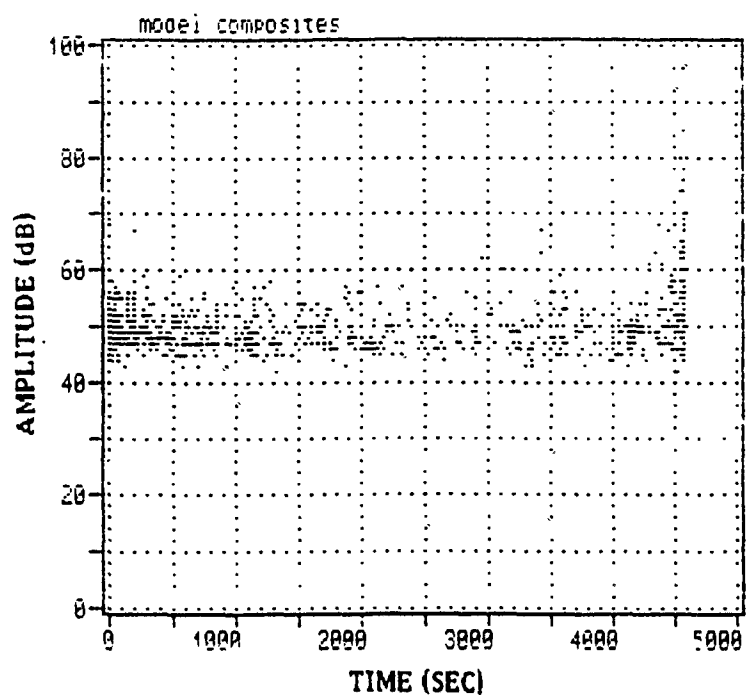


FIGURE A3f. AMPLITUDE vs TIME

1991 USAF SUMMER FACULTY RESEARCH PROGRAM/

GRADUATE STUDENT RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Research & Development Laboratories

FINAL REPORT

Depley of Laminated Panels with Perforation due to impact

Prepared by:	John Lair and David Hui
Academic Rank:	Graduate Student and Associate Professor
Department and	Mechanical Engineering
University:	University of New Orleans
Research Location:	WRDC/FIVST
USAF Researcher:	Greg Czarnecki
Date:	September 29, 1991

Deply of Laminated Panels with Perforation due to Impact

by

John Lair and David Hui

ABSTRACT

Deply techniques were used in composite material laminated plates were used in low velocity impact of panels where there no perforation occurs. Some preliminary data were obtained by Foos (1989). Such deply technique by pyrolysis (heat it in an oven at approx 800 deg F for about 40 min and then separate the plies with a scalpel) has not been commonly applied on perforated plates. Such deply technique will give more accurate delaminated areas and hole sizes than C-Scan techniques. Further, matrix cracks were observer in deply laminates which cannot be observed by C-Scan. Six panels from previous composite panels perforated by spherical projectiles (Altamirano 1991) are used for deply and the separated plies are analyzed in an image processing machine. The delaminated areas are seen quite clearly since they were pre-soaked in gold chloride prior to heating in an oven.

1. INTRODUCTION

The purpose of the work is to determine whether more information can be obtained from deply techniques than from the traditional C-Scan photos. The panels were soaked in gold chloride liquid prior to heating in an oven so that it will leave a gold imprint on the plies which can be easily visualized. The size and shape of the delaminated area as a function of the plies were observed and measured using an image processing machine.

Figure 1a,b shows typical actual size photos of two adjacent plies from the 365 ft./sec. test taken from the

ACKNOWLEDGEMENTS

The authors would like to thank the guidance of Dr. Arnold Mayer and the assistance of Greg Czarnecki due the course of the research. The literature search on the subject of damage of composite panels due to projectile impact was performed at Wright Patterson Air Force Base library and the assistance of the library staff is gratefully acknowledged. The encouragements of James Hodges, Marty Lentz, Eric Fletcher and Pat Petit were appreciated.

Altamirano (1991) test program. The boundaries of the delamination zone were marked with white paint. Such painting was not necessary in the image processing machine since the gold imprints were clearly observable, they are shown only for xeroxing purposes. The matrix cracking can be seen in ply 14 of Fig. 1 and a often occurring "H" shaped deply area can be seen in ply 13. These two plies were adjacent to one another ply 14 being closer to the impact face. The side of the ply closest to the impact face is shown in both cases.

Figure 2a-2f shows graphs of the delamination areas and hole sizes. It can be seen that there exist drastic changes in delamination areas from one ply to the next ply. The hole sizes for relatively low impact velocities (365 ft/sec) would be smaller than that from higher velocities (879, 1013, 2777, 4476, 5817 ft/sec). On the other hand, the delamination areas would be larger for smaller velocities since one can expect that most of the damage would occur at velocities just above the V50.

Figure 3 shows a graph of the delamination volume and hole volume versus the energy absorption of the panels divided by the mass of the 1/2 inch diameter spherical steel projectile, using the deply technique. The energy absorption is defined as the difference between the initial and residual kinetic energies of the projectile. It can be seen that the delamination volume decreases for the first three lower velocities and then increases for the remaining three higher velocities. The hole volume tends to reach an asymptotic value for higher energy absorption. Figure 4 presents the same data taken from pixel count of C-Scan image.

Figure 5 shows a typical C-Scan photo of the entire laminated panel. The pixel count on C-Scan would enable one

to obtain the delamination volume and hole volumes. The delamination volumes measured from the C-Scan image were found to be larger than that obtained from deply technique. This is due to the fact that C-Scan can not detect delaminations which are smaller than those found near the surface. If a delamination overlays a undelaminated area, such undelaminated area is counted as delaminated area, thus overestimating the undelaminated area. On the other hand, the hole size was smaller on the C-Scan results than from the deply results. As these observations are based on the deply of only seven panels, they need to be substantiated by a enlargement of the test program.

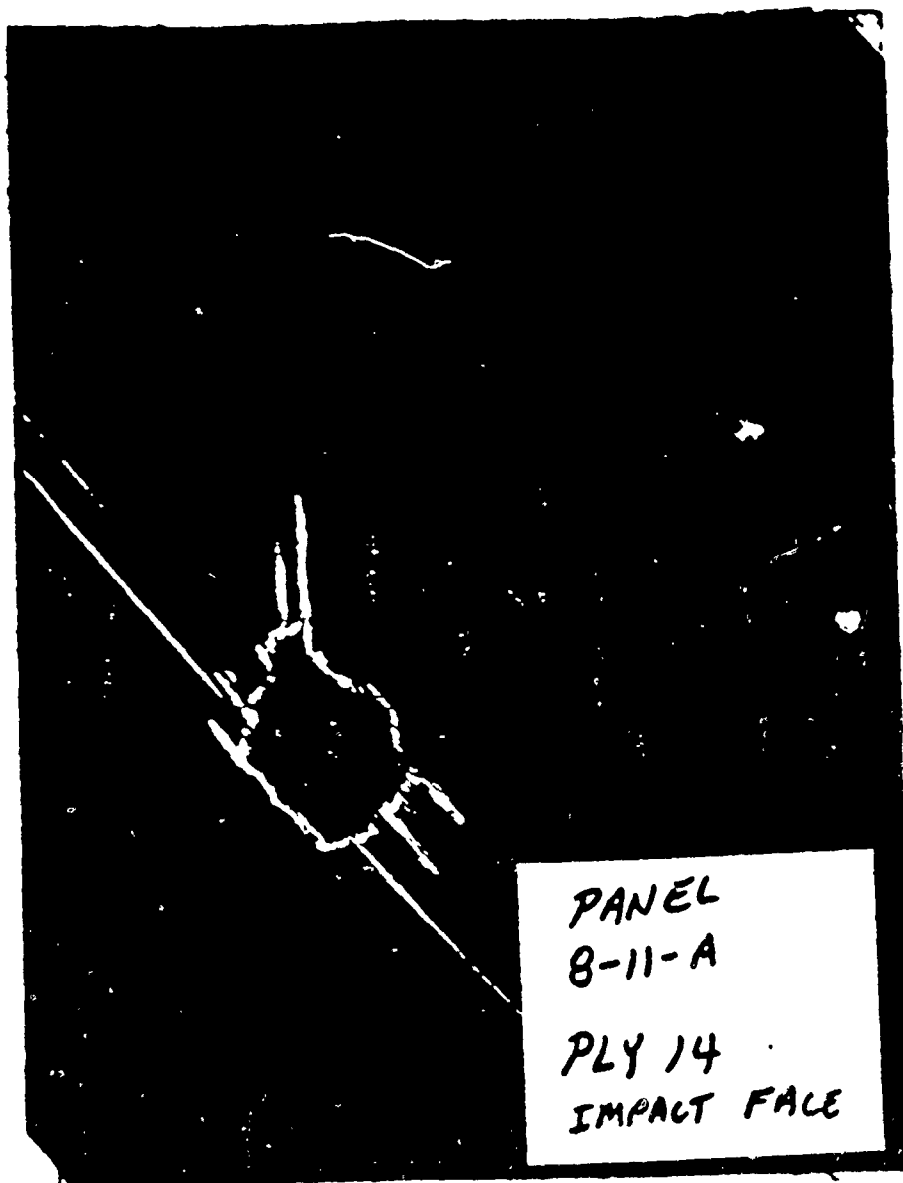
3. CONCLUSIONS

The deply technique was successfully employed to study the damage zone of laminated panels perforated by spherical indenter. Such deply technique was found to give valuable information on the size, shape and matrix cracks which could not be accurately observed by C-Scan photos.

REFERENCES

Altamirano, Magna, Experimental Investigation of High and Low Impact Energy Absorption of AS4/3502 Graphite/Epoxy Panels, M.S. Degree Thesis, University of New Orleans, New Orleans, LA, May 1991

Foos, Brian, Low Velocity Impact Damage Investigations of Composite Plates in a Drop Tower, M.S. Degree Thesis, The Ohio State University, Columbus, OH, March 1990



PANEL
8-11-A
PLY 14
IMPACT FACE

Figure 1a deplied lamina



Figure 1b deplined lamina

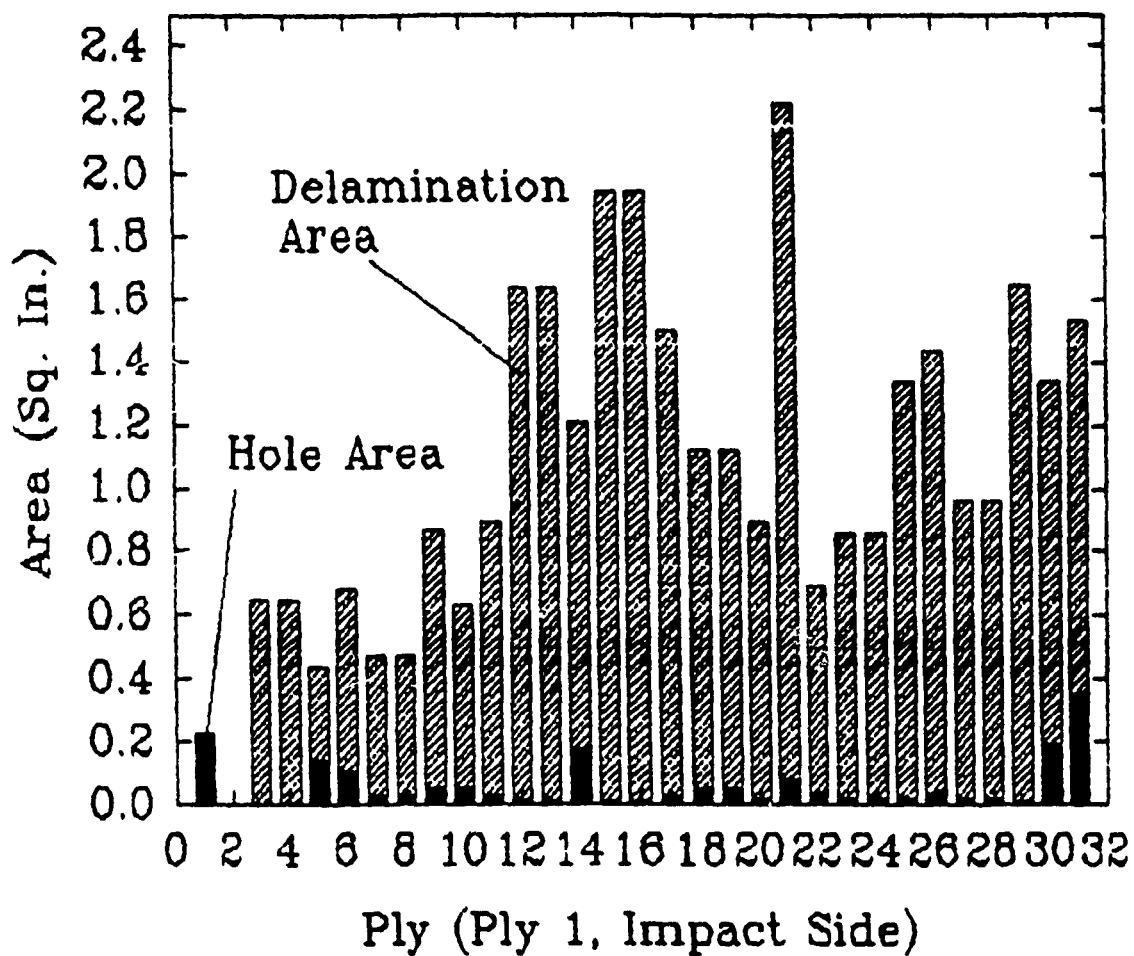


FIGURE 2a

Damage at 365 ft./sec. impact velocity
From Deply technique

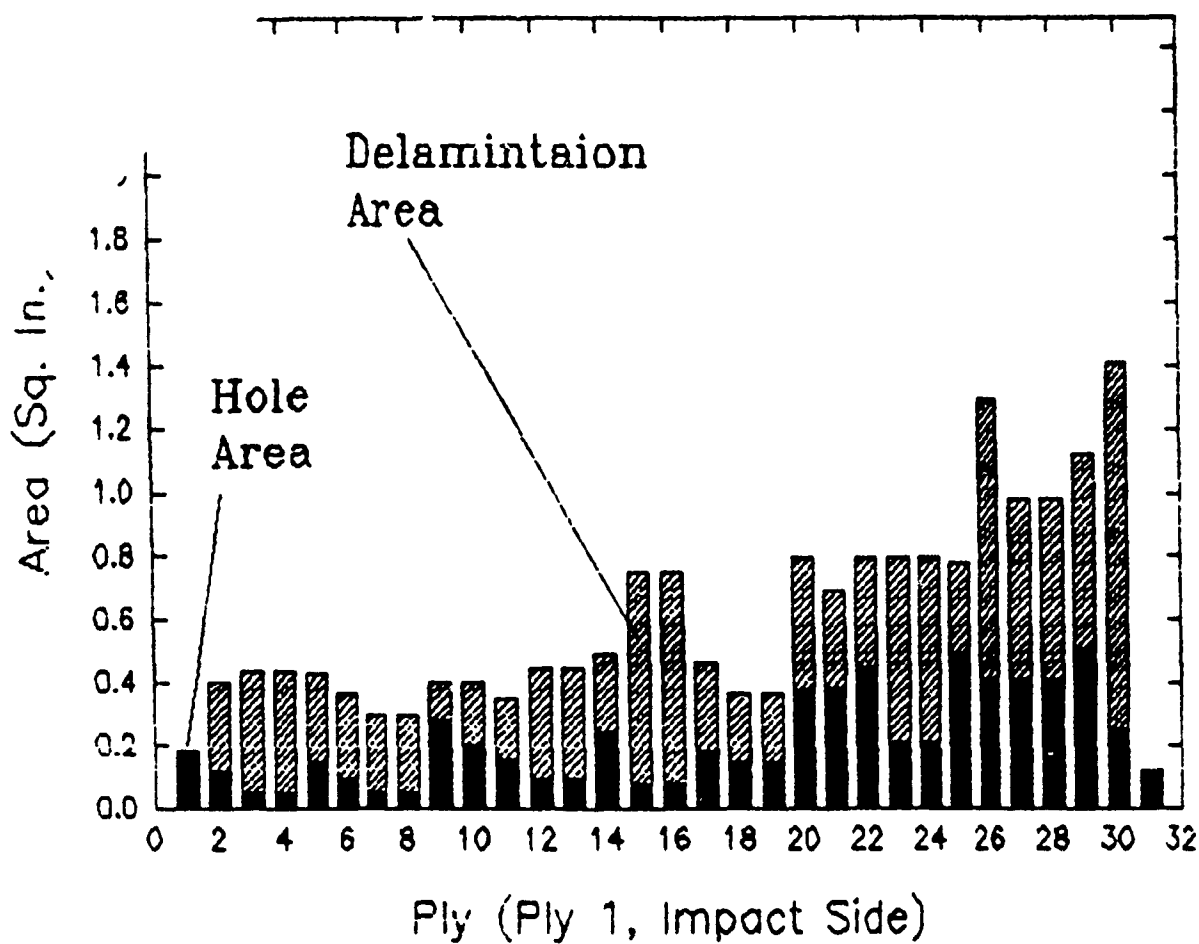


FIGURE 2b Damage at 879 Ft./Sec Impact Velocity.
From Deply Technique

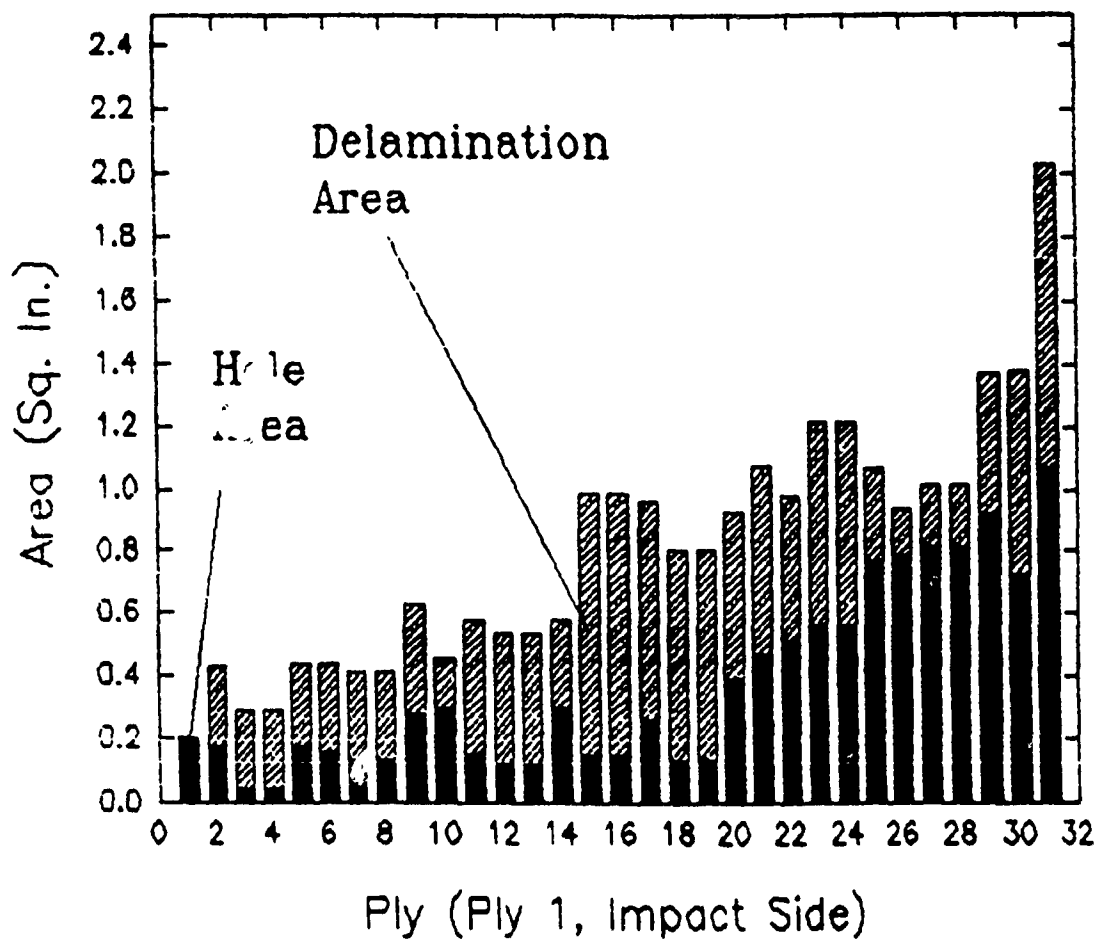


FIGURE 2c Damage at 1013 Ft./Sec. Impact Velocity
From Deply Technique

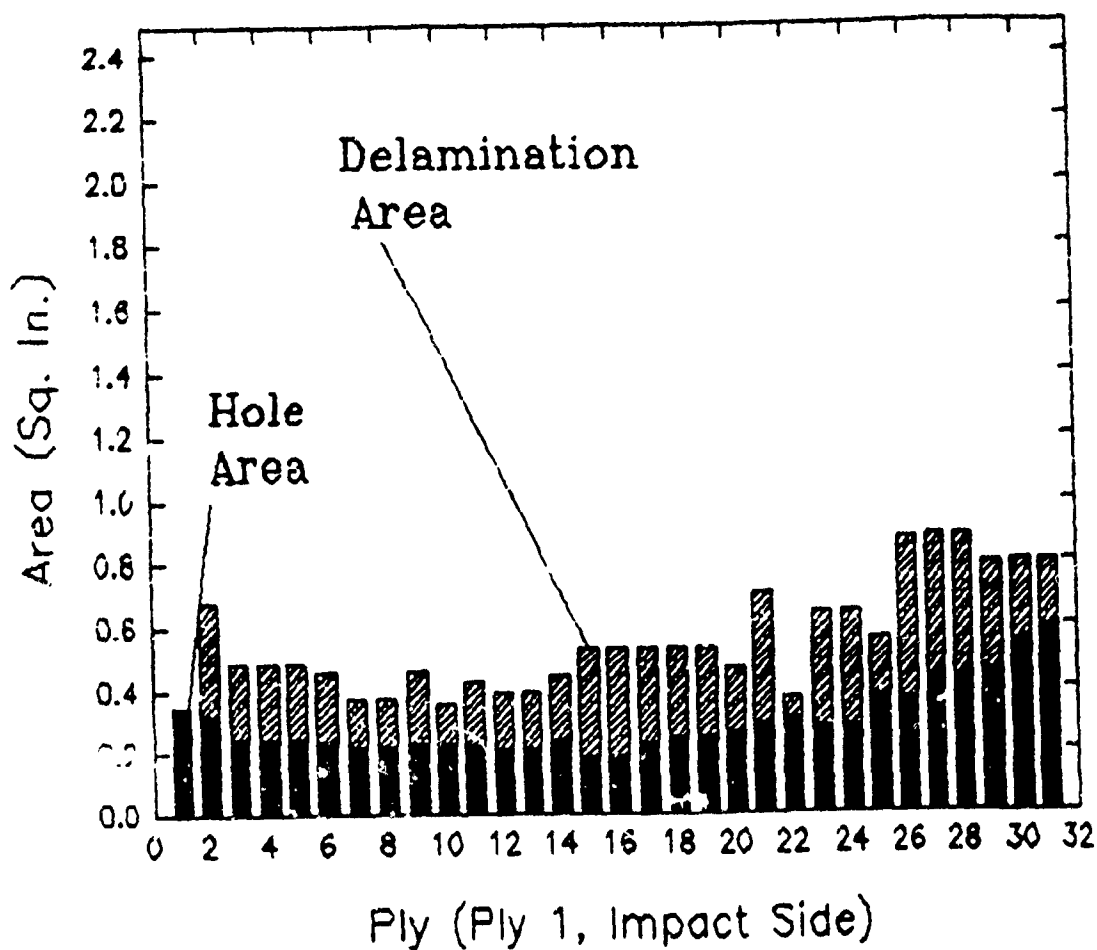


FIGURE 2d Damage at 2777 Ft./Sec. Impact Velocity
From Derv Technique

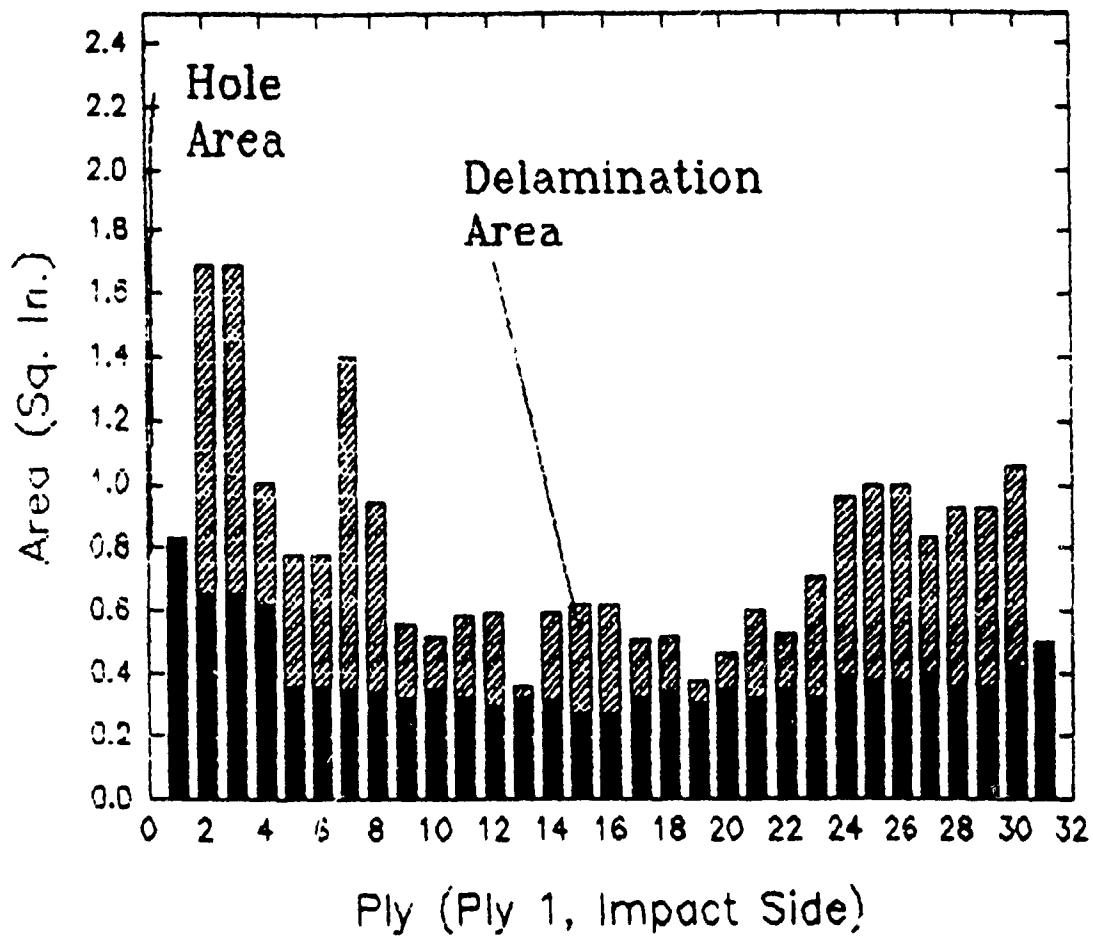


FIGURE 2e Damage at 4476 Ft./Sec. Impact Velocity
From Deploy Technique

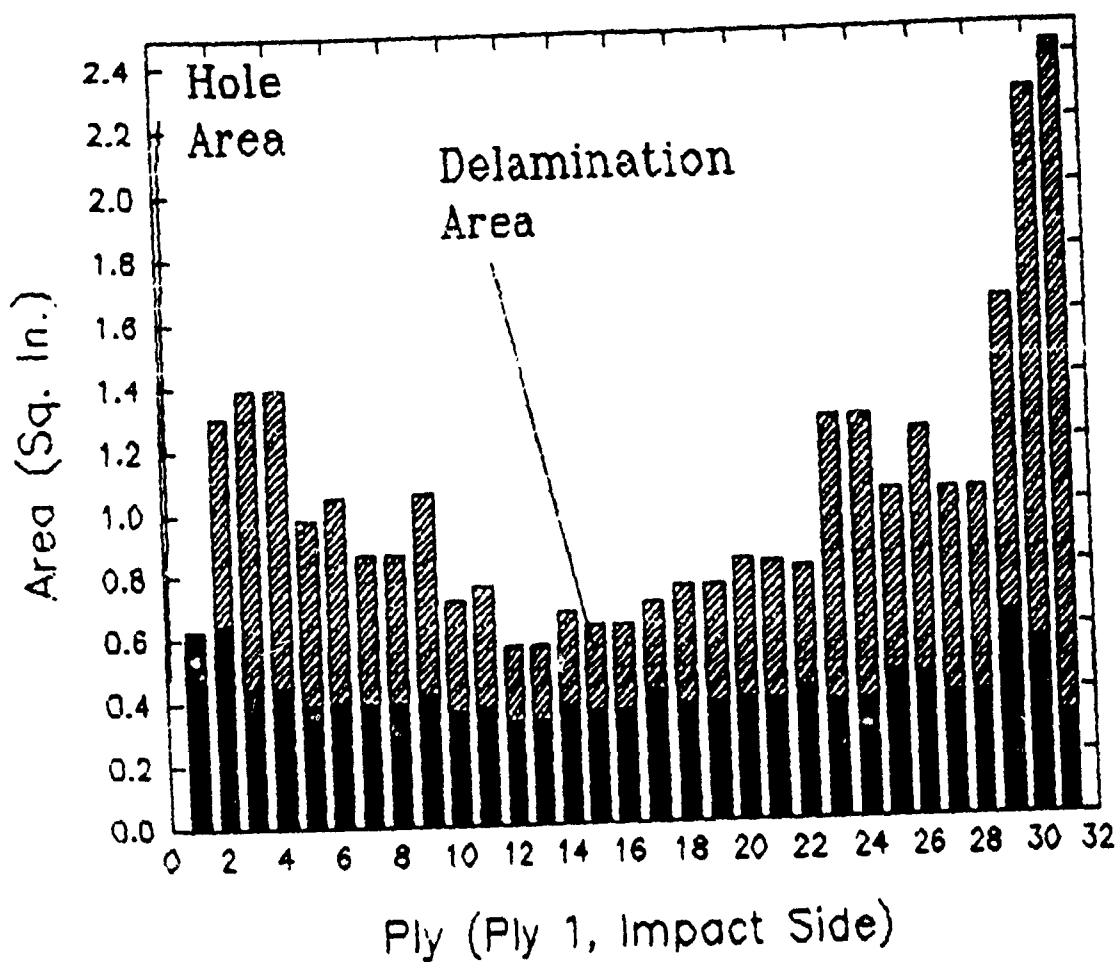


FIGURE 2f Damage at 5817 Ft./Sec. Impact Velocity
From Deply Technique

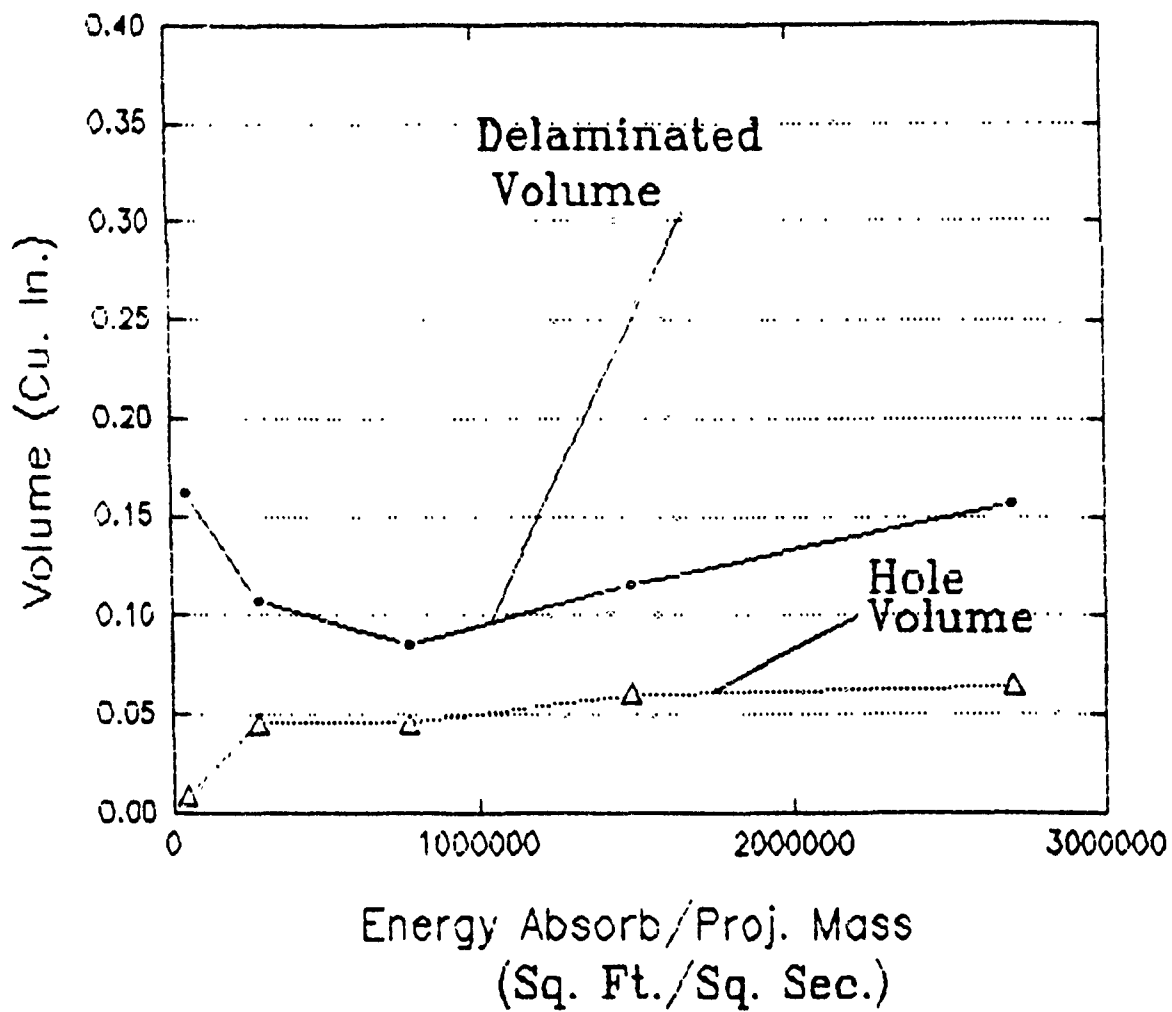


FIGURE 3 Projected Volume of Plug = 0.0314 Cu. In.

Damage Volume Vs Energy Absorbition
Using Deply Technique

SFL MODEL 1750 MINI-C

HRES: 0.020 in.

VRES: 0.020 in.

PEAK STORE: OFF

MODE: TOF GATE: 1

MAGNIFICATION: 1

X POS: in.

Y POS: in.

07/16/91

07:41:29

08:21:50

CSCAN INSPECTION

CLEAR

CALIBRATE

ORIGIN

REFERENCE POINTS

BOUNDARY POINTS

MAG LEVEL

AMP, TOF

GATE SELECT

PEAK STORE

INSPECT

COLOR/GRAY

CONTRAST

HORIZONTAL SLICE

VERTICAL SLICE

READBACK

STORE

DISPLAY NOTES

PROJECTED VIEW

	0.004	0.082
	0.014	0.092
	0.024	0.102
	0.033	0.112
	0.043	0.122
	0.053	0.131
	0.063	0.141
	0.073	0.151
	DEPTH (in.)	

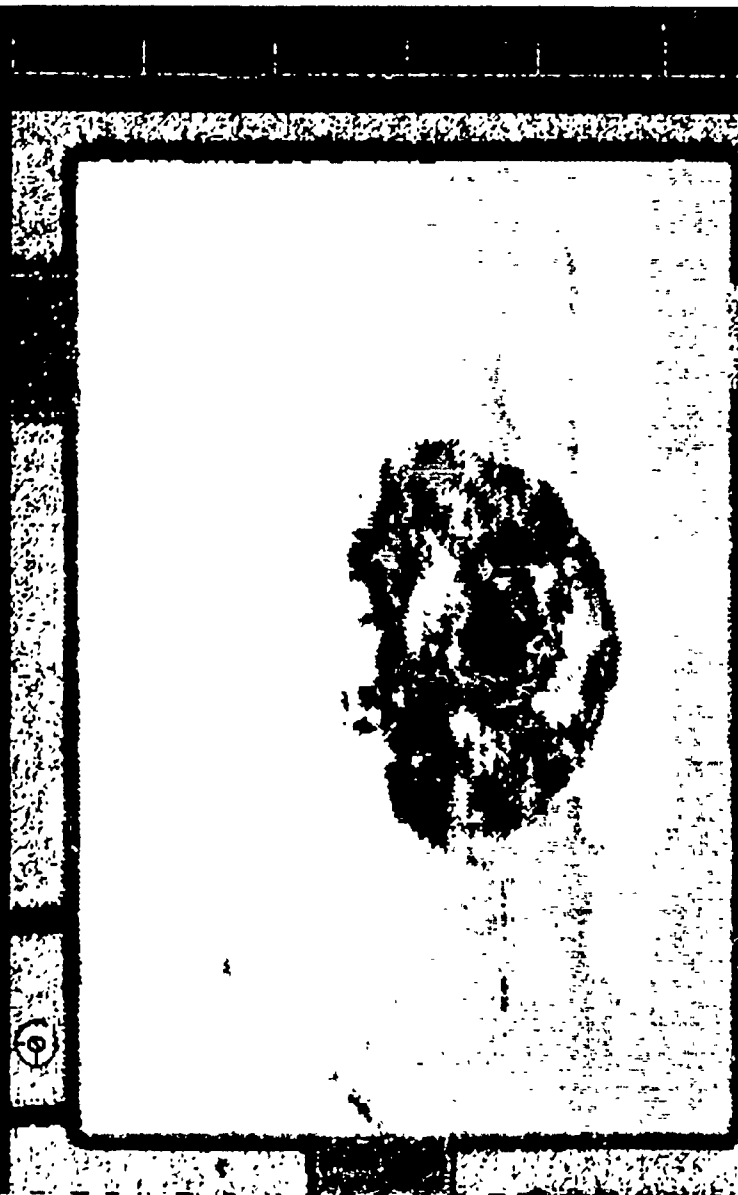


FIGURE 5 C-Scan image

ACTUATOR LOCATION AND OPTIMAL CONTROL DESIGN FOR FLEXIBLE STRUCTURES

John A. Mitchell
Research Associate

ABSTRACT

The optimal placement of one actuator and design of the associated optimal feedback control law is discussed and analyzed for flexible structures using the linear quadratic regulator formulation. This approach is then used to place multiple actuators sub-optimally. An iterative optimization scheme finds the global solution to this problem, and is based upon the theory and algorithm put forth by (Geromel,1989). Application of the algorithm to a small sample problem is presented.

INTRODUCTION

This paper is concerned with optimal control problems using the linear quadratic regulator formulation. Traditionally, this type of problem involves finding the optimal input to the system such that various objectives are achieved i.e. to minimize input to the system while maintaining the state vector close to some prespecified state. The state vector follows according to its governing equation and acts as a constraint in the optimization problem while the corresponding system parameters are fixed. Finding the solution to this problem requires the solution to the Riccati

equation. Given the solution to the Riccati equation the optimal control vector is readily found. Naturally this solution depends upon the system parameters i.e. structural mass, stiffness, damping and also actuator locations. In this paper, the main concern is with actuator locations and optimal input only, and therefore it is assumed that the structural properties are fixed. The corresponding problem is to solve the traditional linear quadratic regulator problem while simultaneously selecting actuator locations to best achieve the desired objectives.

PRELIMINARIES

For flexible structures, the time invariant equation of motion can be expressed in state space form as

$$\dot{\vec{x}} = A\vec{x} + B\vec{u}$$

where A contains stiffness, mass and damping information. B contains the actuator locations, and \vec{x} and \vec{u} are the system states and input vectors respectively. Given some transient excitation, the state \vec{x} is governed by the eigenvalues of A . If this response is undesirable, then an optimal input vector \vec{u} is designed to drive \vec{x} to the desired state. This design is found by defining the linear quadratic cost function as

$$J = \frac{1}{2} \int_0^T (\vec{x}^t Q \vec{x} + \vec{u}^t R \vec{u}) dt$$

where Q and R are the typical weighting matrices associated with the states and inputs respectively. The optimal input \vec{u} can be found to be $\vec{u} = -R^{-1}B^t K \vec{x}$ where

K satisfies the following Riccati equation.

$$\dot{K} = -KA - A^t K - Q + KBR^{-1}B^t K$$

It is obvious that the optimal input depends explicitly on the input matrix B.

In order to implement the above control law, the full state vector must be known and this involves the estimation problem. For this paper it is assumed that the state vector is completely known and available for state feedback. Because of the well known separation principle this is not a restrictive assumption. The eigenvalues of the controller and estimator can be designed separately and the resulting closed loop eigenvalues are the sum of the two sets of eigenvalues.

In the traditional formulation, B is fixed during the optimization. However, in the control of flexible structures, the selection of B can have a significant effect on the final system performance. Therefore the problem is posed so as to let $B = B(\pi)$ depend upon an additional parameter π that may be designed so as to better achieve the objectives. π will be used later to differentiate between particular actuator locations.

In this case the optimal solution to the problem results from the minimization of the cost function J . It is known that the minimal value for J is given by

$$J_{min} = \frac{1}{2} \bar{x}_o^t K_o(\pi) \bar{x}_o$$

where the explicit dependence of π on K is included in the notation. Also note that the subscripts 'o' denote the initial value of the subscripted variable at time $t = 0$. However, this is an undesirable situation because the minimal value of J depends

upon the initial disturbance to the system. This problem can be circumvented by choosing an appropriate measure $\sigma(\pi)$ associated to $K(\pi)$. The following measures (Geromel,1989), have been shown to have favorable mathematical properties (to be defined later) associated with the optimization problem.

$$\sigma_M(\pi) = \max_{\|\bar{x}_o\|=1} J(\pi, \bar{x}_o) = \frac{1}{2} \Lambda_M\{K(\pi)\}$$

$$\sigma_o(\pi) = E\{J(\pi, \bar{x}_o)\} = \frac{1}{2} Tr\{K(\pi)\Xi\}$$

In the above equations E is the mathematical expectation operator, $Tr\{*\}$ is the trace of $\{*\}$, Λ_M is the maximum eigenvalue of $K(\pi)$, and Ξ is the covariance matrix of \bar{x}_o . Using one of the above measures, the following optimization problem is defined:

$$optimize\{\sigma(\pi), \pi \in \phi\}$$

where ϕ is a set of design constraints.

ACTUATOR LOCATION AND CONTROL DESIGN

The purpose of this section is to provide the necessary equations and algorithm to perform the optimization defined in the previous section. For a more detailed description of the theoretical aspects of this procedure see (Geromel,1989).

The objective function described in the previous section is used here.

$$J = \frac{1}{2} \int_0^T (\bar{x}^T Q \bar{x} + \bar{u}^T R \bar{u}) dt$$

For clarity, it is useful to rewrite the constraint equation with the explicit dependence of the actuator locations included:

$$\dot{\bar{x}} = A\bar{x} + B(\pi)\bar{u}$$

where B and π are defined as follows:

$$B(\pi) = [B_1\pi_1 \ B_2\pi_2 \ \dots \ B_N\pi_N]$$

$$\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_N]$$

The vector $\pi \in \mathbb{R}^N$ is composed of N elements $\pi_j \in \{0,1\}$ such that if $\pi_j = 0$, there is no actuator present and if $\pi_j = 1$, there is an actuator present. The purpose here is to place one actuator optimally out of a set of N possible locations while simultaneously designing the optimal control law. More generally, this may be visualized as selecting an optimal set of column(s) B_j for $j = 1, 2, \dots, N$ in conjunction with the design of the optimal control law. The traditional optimal control formulation requires the pair $\{A, B\}$ be controllable. Therefore define the set ϕ where $\pi \in \phi$ such that for all π the pair $\{A, B(\pi)\}$ is controllable.

The essence of the approach suggested by Geromel is a subgradient calculation of $\sigma(\pi)$. For any $\pi \in \phi$, the subgradient μ is defined as follows:

$$\mu(\pi) = [\mu_1(\pi) \ \mu_2(\pi) \ \dots \ \mu_N(\pi)] = Tr\{L_j S(\pi)\}$$

where $L_j = B_j R_j^{-1} B_j^t$ for $j = 1, 2, \dots, N$. For the infinite time horizon case, i.e. when T is taken as $T = \infty$ in the cost function J , the steady state solution $K(\pi)$ of the Riccati equation is taken, and $S(\pi)$ is defined as follows:

$$S(\pi) = -\frac{1}{2} K(\pi) \theta(\pi) K(\pi)$$

where $\theta(\pi)$ is the solution to the Lyapunov equation

$$[A - B(\pi)G]\theta(\pi) + \theta(\pi)[A - B(\pi)G]^t\theta(\pi) + Z = 0$$

and Z is an arbitrary positive definite matrix chosen such that a positive definite solution $\theta(\pi)$ exists. The matrix $G = R^{-1}B^t(\pi)K(\pi)$ corresponds to the optimal feedback gains.

Before proceeding with the optimization algorithm, it will be helpful to describe a mathematical relationship between ϕ and $\sigma(\pi)$. It has been demonstrated (Geromel, 1989) that both $\sigma_M(\pi)$ and $\sigma_o(\pi)$ are convex over the set ϕ_c where ϕ is a subset of ϕ_c . This convexity property allows for the global optimization of the actuator location/ control problem and is of fundamental importance to an understanding of the theoretical basis of this paper.

Using the following decomposition scheme, the global solution to the optimization problem may be found.

STEP 1: Let $\pi^o \in \phi$. Select an appropriate measure σ_M or σ_o . Solve the Riccati equation and calculate the subgradient μ^o as defined above. Set $k = 0$ and choose $\epsilon > 0$ sufficiently small. For discrete problems in the spatial domain, numerical experience suggests that ϵ may be chosen arbitrarily close to zero.

STEP 2: Calculate α^{k+1}

$$\alpha^{k+1} = \min_{1 \leq j \leq N} \max_{0 \leq i \leq k} \{\sigma(\pi^i) - Tr\{\mu^i \pi^i\} + \mu_j(\pi^i)\}$$

If j^* is the optimal index, then $\pi_j^{k+1} = 1$ for $j = j^*$ and for $\pi_j = 0$ for $j \neq j^*$.

STEP 3: Solve the Riccati equation and obtain $\sigma(\pi^{k+1})$. If $\sigma(\pi^{k+1}) - \alpha^{k+1} \leq \epsilon$ stop.

Otherwise determine μ^{k+1} , increase k by one and return to STEP 2.

EXAMPLE PROBLEM

In the previous section, a procedure was outlined to select an optimal actuator location from a given set of possible locations. However, for practical applications, it is typically necessary to place more than one actuator. This section describes how the placement of one actuator at a time may be used to place multiple actuators sub-optimally. Results of an example problem are used to demonstrate the effectiveness of the approach.

The basic idea is to begin by placing one actuator, which leaves a possible $N - 1$ locations where a second actuator can be placed. As stated earlier, placing one actuator is the same as selecting an optimal column B_j from a set of $j = 1, 2, \dots, N$ possible columns. It is convenient to define the following matrix P :

$$P(\pi) = [P_1 \ P_2 \ \dots \ P_N]$$

which is composed of the set of possible columns, one of which may be selected as the optimal column. After placing one actuator i.e. selecting an optimal column from the above matrix P , the B and P matrices may be redefined as follows:

$$B(\pi) = [B_1\pi_1 \ B_2\pi_2 \ \dots \ B_{N-1}\pi_{N-1}]$$

$$P(\pi) = [P_1 \ P_2 \ \dots \ P_{N-1}]$$

where a new $B_j = [B_{01} \ P_j]$ is defined for $j = 1, 2, \dots, N - 1$ and B_{01} is the optimal column selected from the original P . In addition, the new P has been defined

by removing the optimal column B_{01} from the original P . This may be repeated indefinitely until the number of actuators requiring placement have been placed. For example, after one actuator has been placed, the new $B(\pi)$ will be an $n \times 2(N-1)$ matrix. If after placing i actuators it is desired to place another, the new $B(\pi)$ will be an $n \times (i+1)(N-i)$ matrix, where there are $N-i$ possible configurations with each new configuration having $i+1$ actuators. Of course after the optimal π has been found, there will only be $i+1$ non-zero columns in B .

The effectiveness of the procedure is good for small order systems as the example will demonstrate. However, as the order of the system increases, there is a decreasing chance for the optimal location of one set of actuator locations to be imbedded in a larger set of optimal locations. To demonstrate the efficacy of the algorithm, a discrete system consisting of masses, springs and dashpots connected in series has been used. There are a total number of 16 states in the state space format and $N = 8$ possible actuator locations. For this problem, it is desired to place two actuators. The equation of motion in state space form is used where

$$A = \left[\begin{array}{c|c} -m^{-1}c & -m^{-1}k \\ \hline I & 0 \end{array} \right] \quad B(\pi) = \left[\begin{array}{c} m^{-1}b(\pi) \\ 0 \end{array} \right]$$

The stiffness, mass and damping matrices are given by m , k , and c respectively. The possible actuator locations are defined in the matrix b which for this problem is the identity matrix of dimension 8. This is equivalent to saying that each mass is a potential location for an actuator. For a definition of specific system parameters see the appendix.

Using the algorithm discussed here, the first actuator was placed optimally for

$\pi^0 = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$. The optimal π was found to be $\pi^* = [0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$.

Table 1 below lists the 8 possible solutions for the problem and the associated cost.

The numbers listed as the cost in Table 1 and in subsequent tables were found through integration of the cost function J over a period of time $t = 0$ to $t = T$. The final time T was determined by monitoring the change of J over each time step.

The same initial conditions were used for all cases.

Cost	π
174.5	1 0 0 0 0 0 0 0
150.7	0 1 0 0 0 0 0 0
135.7	0 0 1 0 0 0 0 0
108.8	0 0 0 1 0 0 0 0
107.1	0 0 0 0 1 0 0 0
100.9	0 0 0 0 0 1 0 0
102.0	0 0 0 0 0 0 1 0
103.2	0 0 0 0 0 0 0 1

Table 1: First Actuator

Proceeding with the sub-optimal approach for the placement of the second actuator, as described previously in this section, the second optimal location was found to be $\pi^* = [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$. Table 2 below lists the cost associated with each possible solution for the placement of the second actuator. After combining the two solutions, the final configuration found was $\pi = [0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0]$.

Now the question remains as to how optimal the final solution is in terms of the global minimum. For $N = 8$, there is a total of 27 possible solutions, and Table 3 below lists the minimal cost associated with each. The sub-optimal approach found a sub-optimal solution. However, as Table 3 indicates, the solution found is very

Cost	π
84.13	1 0 0 0 0 0 0
79.88	0 1 0 0 0 0 0
77.19	0 0 1 0 0 0 0
74.80	0 0 0 1 0 0 0
76.75	0 0 0 0 1 0 0
95.94	0 0 0 0 0 1 0
96.04	0 0 0 0 0 0 1

Table 2: Second Actuator

close to the global solution. In fact it is within 1.3% of the global solution.

FUTURE WORK

The placement of actuators in conjunction with designing the optimal control law for a system is a very important problem. This paper has primarily been concerned with placing one actuator at a time optimally and more than one sub-optimally in this context. There are several other closely related problems where additional work may be done.

First, it may be possible to find the global solution to the multiple actuator placement and control problem. In the paper by (Geromel,1989) the ground work was provided to achieve this. The algorithm reported in the present report is a special case of a more general algorithm which allows for a global optimization of the multiple actuator placement problem. The difference between the two algorithms is in Step 2. A modified version of Step 2 (Geromel,1989) is given below:

Step 2: Calculate α^{k+1}

$$\min_{\alpha, \pi \in \phi} \{ \alpha | \alpha \geq \sigma(\pi^i) + Tr(\mu^i, \pi - \pi^i) \} \text{ for } i = 1, 2, \dots, k$$

Cost	π
134.7	1 1 0 0 0 0 0 0
125.7	1 0 1 0 0 0 0 0
97.95	1 0 0 1 0 0 0 0
92.97	1 0 0 0 1 0 0 0
84.13	1 0 0 0 0 1 0 0
84.47	1 0 0 0 0 0 1 0
84.97	1 0 0 0 0 0 0 1
121.5	0 1 1 0 0 0 0 0
95.53	0 1 0 1 0 0 0 0
90.18	0 1 0 0 1 0 0 0
79.38	0 1 0 0 0 1 0 0
80.01	0 1 0 0 0 0 1 0
80.29	0 1 0 0 0 0 0 1
94.72	0 0 1 1 0 0 0 0
89.15	0 0 1 0 1 0 0 0
77.19	0 0 1 0 0 1 0 0
77.16	0 0 1 0 0 0 1 0
77.25	0 0 1 0 0 0 0 1
93.98	0 0 0 1 1 0 0 0
74.80	0 0 0 1 0 1 0 0
74.30	0 0 0 1 0 0 1 0
73.86	0 0 0 1 0 0 0 1
76.75	0 0 0 0 1 1 0 0
76.19	0 0 0 0 1 0 1 0
75.66	0 0 0 0 1 0 0 1
95.94	0 0 0 0 0 1 1 0
96.04	0 0 0 0 0 1 0 1
97.18	0 0 0 0 0 0 1 1

Table 3: Possible Solutions for Two Actuators

This modification requires an optimization routine capable of dealing with complicating variables. The problem is linear and well within the realm of solution realization. An implementation of this algorithm would represent real progress toward an improvement of controlling flexible structures.

One feature that may be utilized for some systems is an imbedding property

of the optimal actuator locations. In the sample problem, there was a significant amount of imbedding present. This is apparent by noticing trends of the cost in Table 3, as actuator locations are systematically varied. In looking at the top two sets of possible solutions, one can see that the optimal in each set contains the solution from part 1 i.e. the first optimal location found. However, this trend does not continue in the following sets although it is very nearly so. Future work to investigate this imbedding property and its presence or lack of presence would be very useful.

Another feature that is not at all present in the algorithm presented here is the number of actuators required for a particular system. Naturally, this falls into the designer domain. However, there may be a point where adding additional actuators to the system may bring back diminishing returns. This feature may possibly be incorporated into the algorithm.

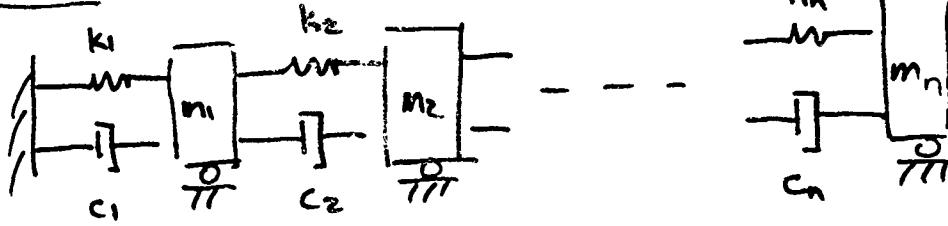
Finally, the problem can be regarded in terms of estimation. Where are the optimal sensor locations? This problem falls within the capability of the algorithm presented here although was not discussed. Implementing an optimal controller and estimator using optimal sensor and actuator locations would yield the ultimate design. This represents another very good area for future work.

REFERENCES

- 1) J. C. Geromel, "Convex Analysis and Global Optimization of Joint Location and Control Problems." *IEEE Trans. Automat. Contr.*, Vol. 34, No. 7, July, 1989.
- 2) H. Kwakernaak, R. Sivan, *Linear Optimal Control Systems*, New York: Wiley, 1972.

APPENDIX

System



for example problem $n=8$

stiffness Matrix $[K]$

$$[K] = \begin{bmatrix} k_1 + k_2 & -k_2 & & & \\ -k_2 & k_2 + k_3 & -k_3 & & \\ & -k_3 & k_3 + k_4 & -k_4 & \\ & & -k_4 & \ddots & \ddots \\ & & & \ddots & \ddots \\ & & & & k_{n-1} + k_n & -k_n \\ & & & & -k_n & k_n \end{bmatrix}$$

where

$$k_1 = 4, k_2 = 12, k_3 = 10, k_4 = 2, k_5 = 6, k_7 = 10, k_8 = 5$$

MASS MATRIX $[M]$

$$[M] = \text{DIAG}[200 \ 175 \ 150 \ 125 \ 100 \ 75 \ 50 \ 25]$$

APPENDIX (continued)

DAMPING MATRIX [C]

damping matrix looks like $[k]$. Replace k_i with c_i for $i = 1, n$.

For example problem $n=8$.

$$c_1 = c_2 = c_3 = c_4 = c_5 = c_6 = c_7 = c_8 = 4$$

FINITE ELEMENT NON-LINEAR MULTIMODE RANDOM RESPONSE OF COMPLEX PANELS

USER'S GUIDE

RUSSELL A. MORRIS
AFOSR SUMMER GRADUATE RESEARCH PROGRAM
VIRGINIA POLYTECHNIC INSTITUTE
SUMMER, 1991

ABSTRACT

An analytical method has been developed for predicting root-mean-squared (RMS) maximum strain and displacements for isotropic and composite plates under random acoustic excitation. Eventually, data from this prediction method may aid in the prediction of sonic fatigue lifetimes of structures such as the National Aerospace Plane (NASP). Work done during the summer has included debugging and updating an already existing finite element code (RBUP2.FOR), generalizing the method of data input to allow for a much broader range of conditions, verifying results with other analytical solutions, performing mesh convergence studies, and developing a method of plotting linear mode shapes for rectangular plates. This paper is intended to serve as a user's guide, discussing in detail changes that were made during the summer.

I. INTRODUCTION

RBUP2.FOR is a finite element program for the prediction of RMS strains and displacements for rectangular composite panels under acoustic excitation. At the time of this report, the package is capable of handling both isotropic and composite plates with any combination of clamped and simply-supported boundary conditions. The following paper has been written to serve as a user's guide to the method which has been developed for creating input files, running the finite element code, and plotting mode shapes. All programs are written in FORTRAN for use on the VAX system. A flowchart of the entire analytical prediction method is shown in Figure 1. below.

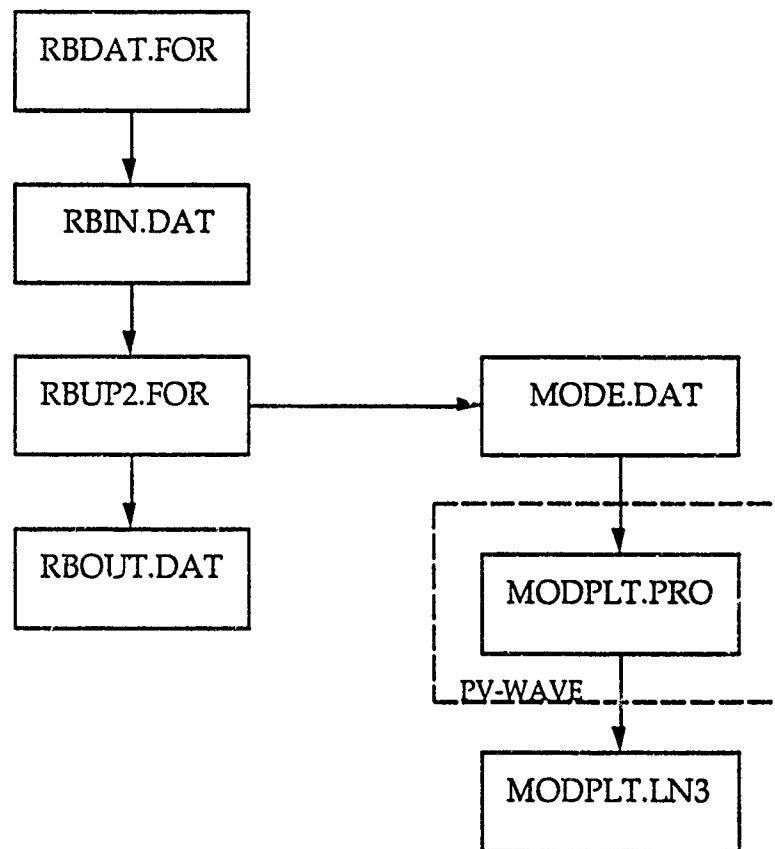


Figure 1. Flowchart for Analytical Package

The prediction method has been broken into three basic steps, each executed separately utilizing three separate FORTRAN programs. The first step is to create the input data file, *RBIN.DAT*. Since the input file contains roughly fifty (50) lines of data, the procedure for creating the file has been

incorporated into the FORTRAN code *RBDAT.FOR*. Program *RBDAT* is an interactive program which prompts users for data input from the keyboard and writes the results in the necessary form to the input file. Once the input file has been created, the second step is the running of the actual finite element code *RBUP2.FOR* which uses the input file *RBIN.DAT* and writes results to the output file *RBOUT.DAT*.

The third step which has been incorporated into the analytical prediction scheme is the plotting of the linear mode shapes. This final step is carried out by a separate FORTRAN program, *MODPLT.PRO*, which can be run from within the plotting program PV-WAVE.

II. INPUT PROGRAM: *RBDAT.FOR*

The input file required by the finite element code is a large file, containing basic data and also a large connectivity matrix. For a plate divided into one hundred elements, the matrix has two hundred and forty components. Calculating a matrix of this size would be very time consuming and typing it to an input file (error free) by hand would be completely unfeasible. The purpose of the interactive input program *RBDAT.FOR* is to relieve the user of the task of typing a new input file for each case of interest. The user is prompted for basic data (material properties, plate dimensions, sound pressure levels, ect...) and boundary conditions. Based on this information, the system degrees of freedom (DOF) are numbered and the connectivity matrix between the global system DOF and the local element DOF is calculated. All information, including the connectivity matrix, is then written to the input data file in the order specified by the finite element code. A listing of *RBDAT.FOR* is located in Appendix A.

A. LABELING SYSTEM DEGREES OF FREEDOM

In order to understand the nature of the connectivity matrix, one must first understand the numbering of the system DOF and the method by which the boundary conditions at the plate edges are applied. The local DOF for a typical plate element are defined as shown in Figure 2. at the top of page 3.

The displacements in the x, y, and z, directions are denoted by u, v, and w respectively. The slopes in the x and y directions are labeled w_x and w_y , while w_{xy} is used to label the twist about the z-axis. The element has a node at each of the four corners, with six DOF defined at each node. Thus there are twenty-four DOF for each element.

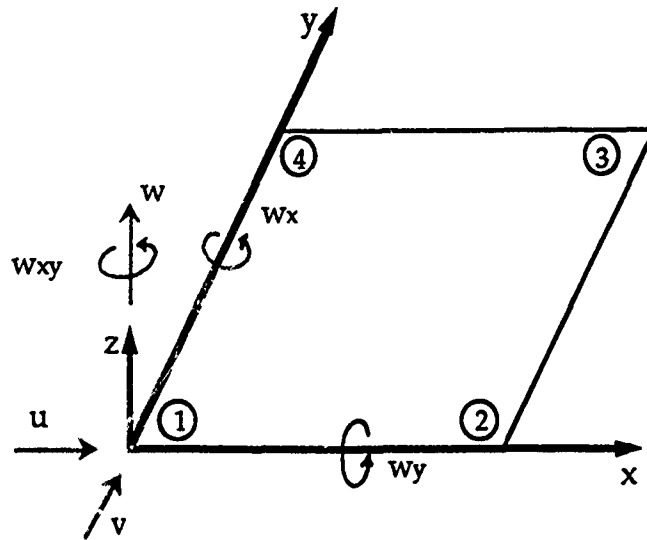


Figure 2. Rectangular Plate Element

A rectangular plate is shown divided into nine elements in Figure 3. The system (global) coordinate system is defined as shown.

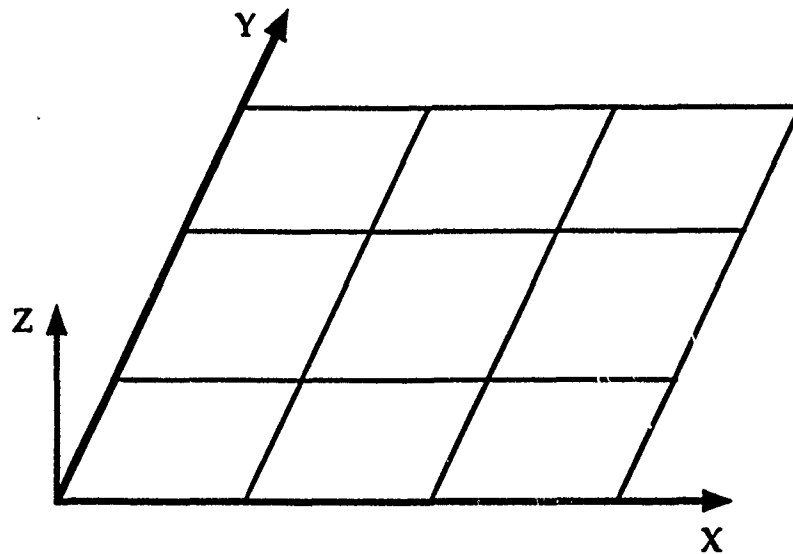


Figure 3. Global Coordinate System

With the elements assembled in this manner, the system DOF are numbered and assembled into the system displacement vector. The displacement vector is partitioned into six sub-vectors containing the displacements in the Z direction, the slopes in the X and Y directions, the

twists about the Z-axis, and the displacements in the X and Y directions respectively, as shown in equation 1.

$$\{\Delta\} = \{(w)(w_x)(w_y)(w_{xy})(u)(v)\}^T \quad (1)$$

The lowest numbered system DOF correspond to the out-of-plane displacements in the Z direction, while the highest numbered DOF correspond to the in-plane displacements in the Y direction.

During the numbering of the system DOF all of the out-of-plane displacements must be numbered first, then the slopes (w_x), and then the rest of the DOF following the order shown in Equation 1. Beginning with the out-of-plane displacement at the node corresponding to the origin of the global coordinate system, the system DOF are numbered moving from left to right along the bottom edge of the plate. When the end of the first row of nodes is reached, then the next highest system degree of freedom is assigned to the node in the second row at the left edge of the plate. The system DOF are numbered in this manner by stepping across the rows in the X direction and progressively moving from bottom to top at the end of each row. When the top right node is reached, then the next highest system degree of freedom is assigned to the slope w_x at the origin and the process is repeated. Following this procedure, all of the system DOF are numbered and ordered by stepping through all of the nodes a total of six times.

B. CONNECTIVITY MATRIX

All of the system DOF are related to the local element DOF through a connectivity matrix. The matrix has twenty-four rows (the number of element DOF) and the number of columns corresponding to the number of plate elements chosen. Each number contained in the matrix is a system degree of freedom. The column and row indicate the element and local degree of freedom to which the system number is related. Note that the matrix is stored in the computer as indicated above, but in the input and output files, the rows and columns are transposed. For example: a section of a connectivity matrix taken from a plate divided into thirty-six (36) elements is shown in Figure 4. at the top of page 5.

0	0	1	0	0	0	6	5	0	13	15	0	0	0	21	0	0	0	25	0	0	0	29	0
0	0	2	1	0	0	7	6	13	14	16	15	0	0	22	21	0	0	26	25	0	0	30	29
0	0	0	2	0	0	8	7	14	0	0	16	0	0	0	22	0	0	0	26	0	0	0	30
0	1	3	0	5	6	10	9	0	13	17	0	0	21	23	0	0	25	27	0	0	29	31	0
1	2	4	3	6	7	11	10	15	16	18	17	21	22	24	23	25	26	28	27	29	30	32	31
2	0	0	4	7	8	12	11	16	0	0	18	22	0	0	24	26	0	0	28	30	0	0	32
0	3	0	0	9	10	0	0	0	17	19	0	0	23	0	0	0	27	0	0	0	31	0	0
3	4	0	0	10	11	0	0	17	18	20	19	23	24	0	0	27	28	0	0	31	32	0	0
4	0	0	0	11	12	0	0	18	0	0	20	24	0	0	0	28	0	0	0	32	0	0	0

Figure 4. Example Connectivity Matrix

The entry "11" in row 5, column 7 indicates that the slope w_x of node 3 on element 5 has been assigned the system degree of freedom 11.

The connectivity matrix is used by the main program when the element stiffness matrices are assembled to form the global stiffness matrix. For each non-zero entry in the matrix, the program adds the specified component of the element matrix to the correct position in the global matrix.

C. BOUNDARY CONDITIONS

The zero elements of the connectivity matrix indicate the nodes at which boundary conditions apply. When the system DOF are numbered, any nodes which are fixed by the boundary conditions are given zeros in the connectivity matrix. When the finite element program reads a zero from the connectivity matrix, then the stiffness from the element matrix corresponding to the zeroed DOF is not added to the system matrix.

With the boundary conditions applied in this manner, the input program is currently capable of handling any combination of simply-supported, clamped, and "cut" boundary conditions. The "cut" boundary condition is useful when symmetry allows for modelling of either a quarter (1/4) or half (1/2) of the plate. The following table shows the DOF which are held at zero for each of the three cases.

Table I. Application of Boundary Conditions

	SIMPLE-SUPPORT	CLAMP	CUT
ZERO	w	w	
DEGREES OF	u	u	
FREEDOM	v	v	
	w_{xy}	w_{xy}	w_{xy}
on edge parallel to y-axis		w_x w_y	u w_x
on edge parallel to x-axis		w_x w_y	v w_y

D. STIFFENERS

At the present time, the input program is capable of incorporating stiffeners into the structure, however, it is only capable of handling stiffeners running in one direction. A typical beam element is shown in Figure 5.

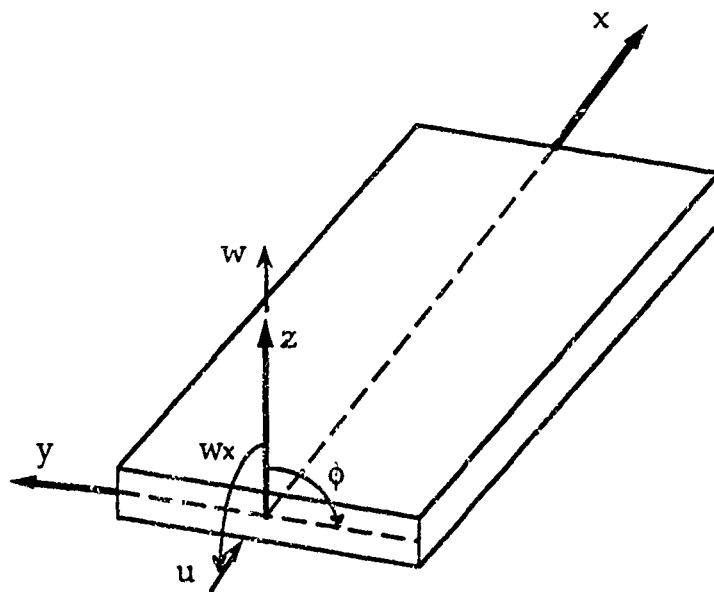


Figure 5. Beam Element

The element has a total of eight DOF (four per node as shown). At each node the displacement in the z direction, slope in the x direction, longitudinal displacement, and longitudinal twist are labeled w , w_x , u , and ϕ respectively.

The stiffeners are placed with the length of the beam running along one side of the plate and divided into a number of elements equal to the number of plate elements along one side of the plate. At common nodes, the corresponding beam and plate element degrees of freedom are equated to insure continuity between the plate and the beam. The beam elements are incorporated into the system matrices through the use of a separate beam connectivity matrix. Once the plate element and system degrees of freedom have been assigned, then the beam element and system DOF relations are found by equating DOF as shown in Table II, and pulling numbers out of the

plate connectivity matrix. Both connectivity matrices are calculated by the input program *RBDAT.FOR*.

Table II. Equivalent Degrees of Freedom

BEAM DEGREE OF FREEDOM	PLATE DEGREE OF FREEDOM
w	w
w _x	w _y
φ	-w _x
u	v

E. RUNNING RBDAT.FOR

In order to run the input program *RBDAT.FOR* on the VAX, the FORTRAN code must first be compiled and linked to create an executable file. The following commands are used:

```
$FORTRAN RBDAT.FOR
$LINK RBDAT
$RUN RBDAT
```

The dollar sign (\$) is the VAX system prompt. Note that once the executable file (*RBDAT.EXE*) has been created, there is no need to compile and link before each run.

The program operates interactively by prompting the user for information which is to be typed into the terminal keyboard. All variables are described on the screen during the run, and in some cases recommended values are given. A sample input file is given in Appendix B.

III. FINITE ELEMENT CODE: RBUP2.FOR

With the implementation of the input file program, running the finite element program has been reduced to a simple task. The input program is the only part of the package which requires any user input. There are, however, a few matrix dimensions which if assigned incorrectly, may adversely effect the results of the program. At the present time the program

has been completely debugged and is currently running with the matrix dimensions discussed in the next section.

A. MATRIX DIMENSIONS

The most common problem encountered during the debugging process involved improper dimensioning of the matrices at the beginning of the main program.

```
COMMON/GAUSS/XA(5),YA(5),WA(5),IG
COMMON/TINV/T(16,16)
COMMON/PARA/ZETA,BWLFR,BWHFR,AL,BL,PI,H,BEL
COMMON/MATRL/AIX,AIZ,G,GJ,P,E,AREA
COMMON/INOUT/IIN,IOUT
COMMON/OTHER/NSDOF
DIMENSION A(3,3),B(3,3),D(3,3)
DIMENSION PEM(24,24),PEK(24,24),PEG(24,24)
DIMENSION BEM(8,8),BEK(8,8),BEG(8,8),TRF(541,100)
DIMENSION SM(541,541),SK(541,541),WK(541,3),KB(24,100)
DIMENSION SKA(100,100),SKB(540,100),SKC(540,540)
DIMENSION RM(100,100),RK(100,100),BV(540),IPIVOT(540)
DIMENSION VAL(100),UMAX(541),SO(541),VR(541)
DIMENSION WMS(541,541),QN(541),WMAX(541),WMAX1(541)

DIMENSION VEC(100,9),AMODEF(541,9),AMODE(541,9),DEL1(24,9)
DIMENSION A1(9),B1(9),C1(9),D1(9),E1(9),F1(9),G1(9),H1(9)
DIMENSION QMS(9),SUMOD(9),DM(9,9),DK(9,9)
DIMENSION R1(9),R2(9),R3(9),R4(9,9),R5(9,9),R6(9,9)

DIMENSION BX2U(2,2,100),BY2U(2,2,100),EXY2U(2,2,100)
DIMENSION DUMMY(100),OMG2(100),DUMP(24),DUMB(8)
DIMENSION KBB(8,100)
open(unit=2,file='rbin.dat',status='old')
OPEN(UNIT=6,FILE='RBOU.DAT',STATUS='NEW')
```

Figure 6. Matrix Dimensions

Figure 6 shows a copy of several lines from the main program. The dimensions shown were established for use with plates divided into one hundred (100) elements or less. Convergence studies have shown that there is probably not a need to divide the plates into a higher number of elements, however, the following section will describe the dimensions in case the need arises.

Most of the matrices shown in Figure 6 may be given dimensions arbitrarily large without having an effect on the results of the program. The lower bound on the dimensions depends on the size of the system (ie. the

number of elements chosen for the plate model). The numbers shown in the figure are explained below:

541: This dimension was chosen to be greater than or equal to the maximum total number of system degrees of freedom. All of the system matrices and vectors are given this dimension.

100: This dimension was chosen to be greater than or equal to the number of reduced system degrees of freedom.

Note: When Guyan Reduction is carried out, the system displacement vector is reduced to hold only the out-of-plane displacements. The number representing the reduced system DOF is equal to the number of system DOF assigned to out-of-plane displacements.

540: This dimension was chosen to be greater than or equal to the difference between the total number of system DOF and the reduced system DOF.

9: This dimension was chosen to be greater than or equal to the number of linear modes included in the non-linear formulation.

In addition to the variable dimensions shown above, there are several matrices that have one or two fixed dimensions which should not be changed. These are listed in Table III below.

Table III. Matrices With Fixed Dimensions

XA(5)	YA(5)	WA(5)	T(16,16)	A(3,3)
B(3,3)	D(3,3)	PEM(24,24)	PEK(24,24)	PEG(24,24)
BEM(8,8)	BEK(8,8)	BEG(8,8)	DUMP(24)	DUMB(8)
WK(_,3)	KB(24,_)	KBB(8,_)	DEL1(24,_)	EX2U(2,2,_)
EY2U(2,2,_)				EXY2U(2,2,_)

Note: A blank space (_) indicates space left for one of the variable dimensions described above and shown in Figure 6.

There is one final matrix which must have the correct dimension in order for the finite element program to run correctly. Matrix IWK is located in Subroutine Reduce. The dimension of IWK must be greater than twice the difference between the number of system DOF and the reduced system DOF. Matrix IWK is a work space for a different subroutine. It is only important that the work space is large enough and it may be arbitrarily large.

```
c      NT = LAST NUMBER OF CURVATURE OF W
c      NSDOF = ACTUAL NUMBER OF SYSTEM DOF
c
      NS = 541
      READ(2,*)NE,NEM
```

Figure 7. RBUP2.FOR Lines 50 - 54

Figure 7 shows lines 50 - 54 of the finite element code. The variable NS defined in line 53 is passed to all the subroutines and used to dimension the system matrices. It is **VERY IMPORTANT** that this number be set equal to the number chosen for the dimensioning of the system matrices at the beginning of the main program (541 in the present example); Otherwise, the program will not run correctly and will give incorrect results.

B. RUNNING RBUP2.FOR

The most efficient method of running the finite element program is to submit the run to the system as a batch job. First, *RBUP2.FOR* must be compiled and linked with two additional subroutines, *MATINV* and *SYMGEF*. This may be accomplished through the use of the .com file *COMPRB.COM*: which consists of the following commands:

```
$FORTRAN RBUP2.FOR
$LINK RBUP2,MATINV,SYMGEF
$EXIT
```

To execute this .com file, simply type: @COMPRB. When the execution is complete, then the executable file RBUP2.EXE will exist.

Once the executable file has been created, it may be run in batch mode by typing **SUBMIT RBUP2.COM**, where *RBUP2.COM* is the .com file consisting of the following commands:

```
$SHOW/PROCESS/ALL  
$RUN RBUP2  
$EXIT
```

The file will run in batch mode and write the output file *RBOUT.DAT* to the user's disk. Running in batch mode allows several cases to be submitted and run at the same time. A Sample output file is located in Appendix C.

IV. PLOTTING MODE SHAPES

When the finite element code is run, a second data file (*MODE.DAT*) is created in addition to the output file *RBOUT.DAT*. *MODE.DAT* contains the linear mode shapes which are a result of the linear analysis.

PV-WAVE is an interactive plotting package in which commands may be entered in line mode or series of commands may be saved as wave programs. This package was chosen for the task of plotting the linear mode shapes because of its programmable nature and the fact that it is readily available on the VAX. When in the **PV-WAVE** environment, drawings are created by a series of line commands. If these commands are written into a wave program, then the drawing can be created by simply entering the command to run the program. This allows users to create plots with very little knowledge of the actual plot commands. *MODPLT.PRO* is a wave program that has been created to read data from *MODE.DAT* and create a surface plot of the rectangular plate showing the out-of-plane displacements. The surface plot is drawn to the screen and a file (*MODPLT.LN3*) is created that may be submitted to the LN03 printer to obtain a hard copy. A listing of *MODPLT.PRO* is given in Appendix D, along with sample mode plots of a 12" x 15" plate divided into one hundred (100) elements.

In order to create plots of mode shapes from *MODE.DAT*, on the LN03 printer, the following steps must be followed:

- 1.) Type **WAVE** (the **WAVE>** prompt replaces the VAX **\$**)
- 2.) Type **.RUN MODPLT** (surface plot appears on the screen)
- 3.) Type **\$PRINT/PASSALL/QUE=TTA6:MODPLT.LN3**

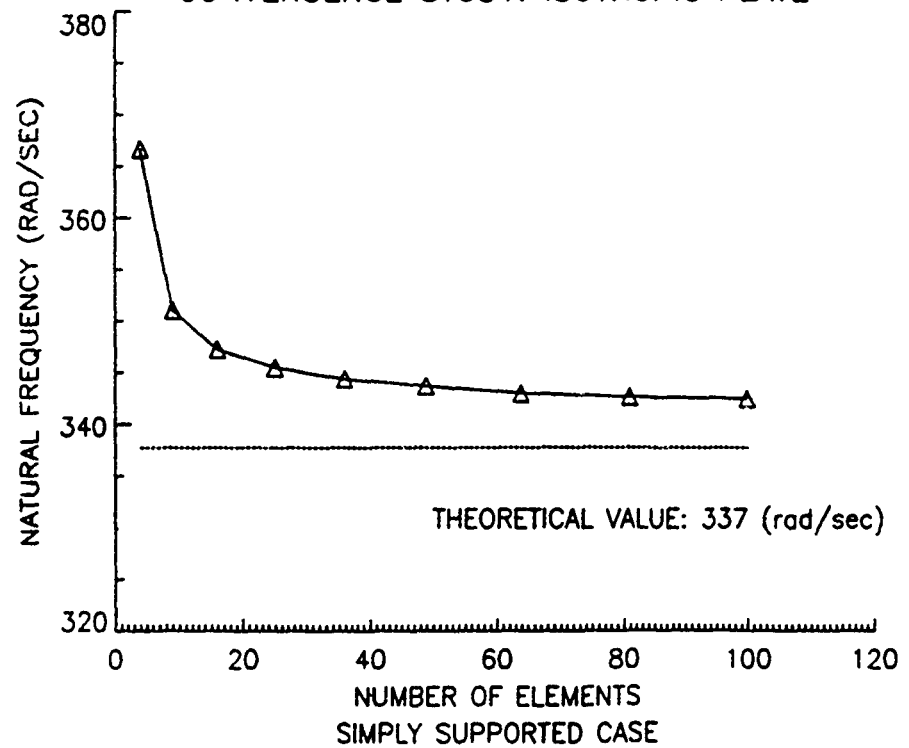
V. CONVERGENCE STUDY:

Finite element codes are usually a trade-off between accuracy and run-time/cost. If the number of elements chosen for the finite element model is very large, then a high amount of accuracy may be obtained in the solution but the code may take an unreasonable amount of time to run to completion. Actually, there is a point for any good finite element code at which the accuracy of the solution changes very little with an increase in the number of elements. In other words, when this point is found the finite element code is giving the most accurate solution it can, given the assumptions made in developing the mathematical formulation. The determination of this optimum number of elements is known as a convergence study.

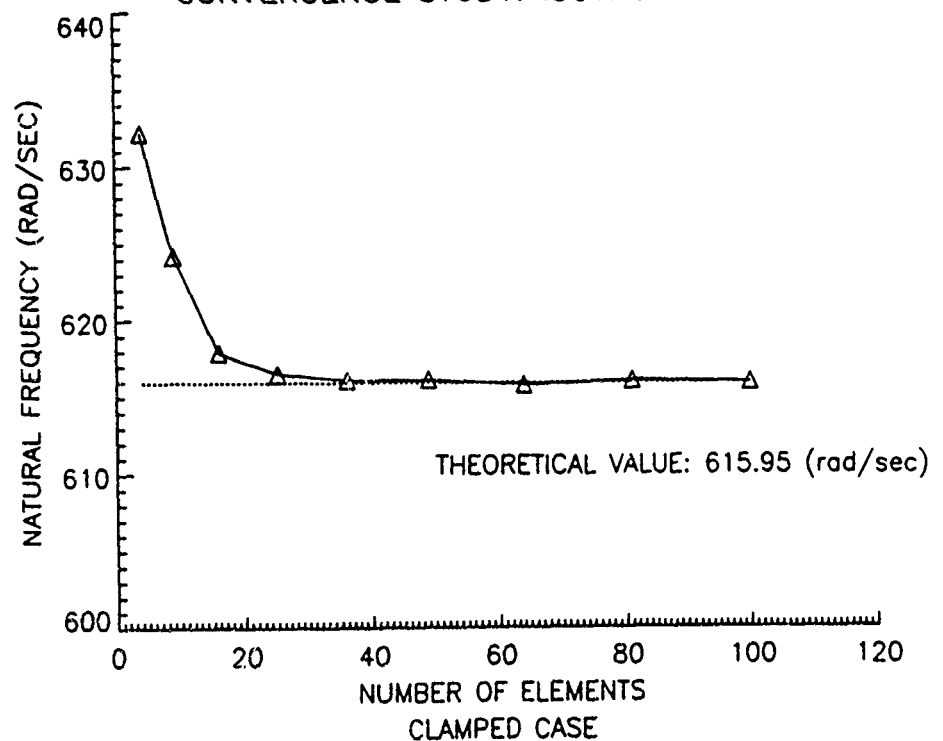
A convergence study was carried out for the finite element code *RBUP2.FOR* to determine the optimum number of plate elements for the analytical prediction of the response of both isotropic (aluminum) and composite (graphite/epoxy) plates. The criterion chosen for the determination of the program accuracy was the natural frequency of plate vibration. Square 12 inch plates were chosen for the analysis. Several cases were run for each set of boundary conditions with the only variable being the number of elements used in the plate model. The result is a plot of natural frequency dependence on the number of elements as shown on pages 13 and 14. From these plots, one can see that the natural frequencies converge rather quickly to within a few percent of the theoretically predicted values. The curves all level off, indicating that there is nothing to be gained (in solution accuracy) from the inclusion of a high number of elements in the plate model. From these curves, one can conclude that the choice of thirty-six elements for the plate model would insure that the solution is within 5% of the theoretically predicted value. All theoretical values were calculated from equations taken from references 1 (isotropic case) and 2 (composite case), based on material properties and plate dimensions given in Appendix E.

It should be noted that further convergence studies will have to be carried out in the future based on the RMS maximum strains and displacements which result from the non-linear analysis. The convergence study on the linear results is meant to serve as a preliminary result, giving a rough estimate of the optimum number of elements, while at the same verifying the fact that the code is running correctly.

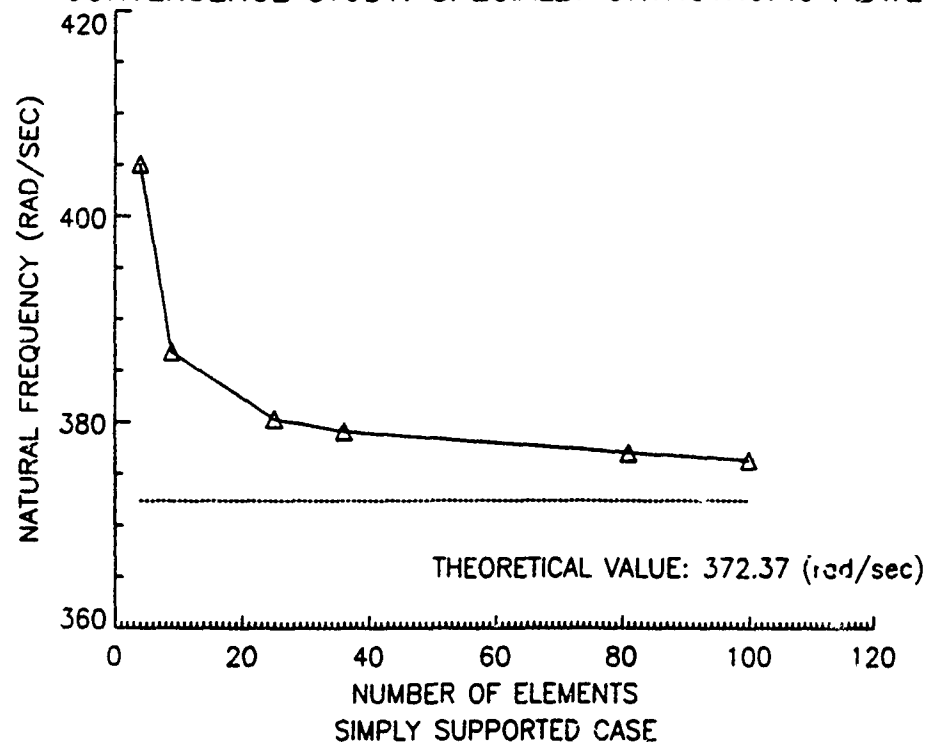
CONVERGENCE STUDY: ISOTROPIC PLATE



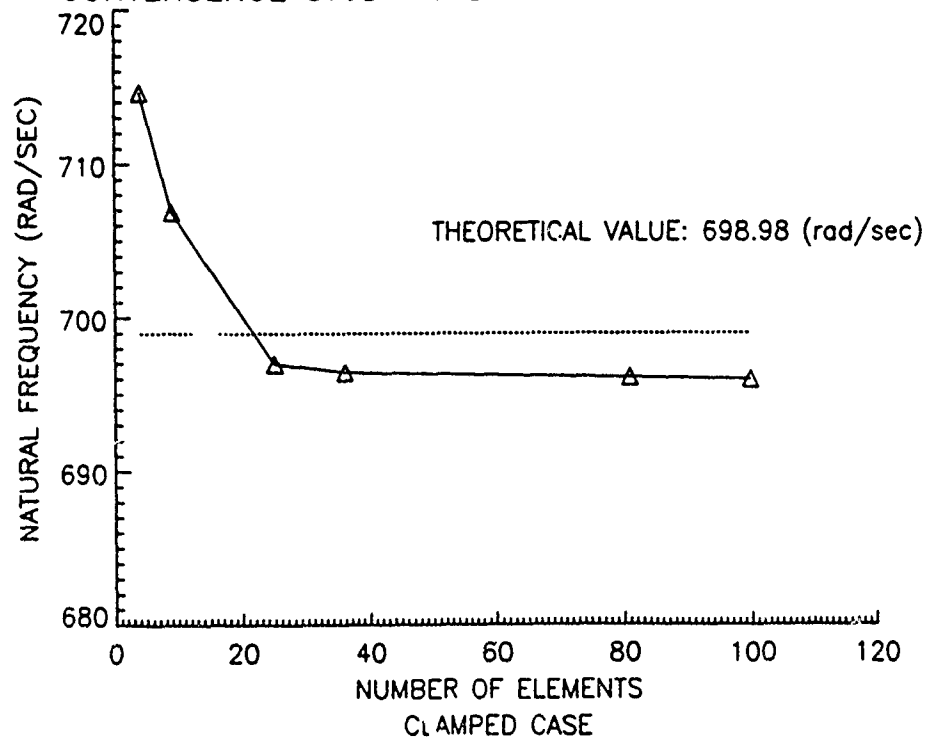
CONVERGENCE STUDY: ISOTROPIC PLATE



CONVERGENCE STUDY: SPECIALLY ORTHOTROPIC PLATE



CONVERGENCE STUDY: SPECIALLY ORTHOTROPIC PLATE



REFERENCES

1. Leissa, Arthur W.. Vibration of Plates. NASA SP-160. Washington D.C.: NASA Scientific & Technical Information Division, 1969, pp.40-50.
2. Ashton, J.E., and Whitney, J.M.. Theory of Laminated Plates. New York: Technomic Publishing Co., 1970, pp. 74-75.

APPENDIX A:
PROGRAM LISTING: RBDAT.FOR

```

C      PROGRAM:      RBDAT.FOR
C      WRITTEN BY:   RUSSELL A. MORRIS
C      WRITTEN FOR:  WL/FIBGD
C
C      UPDATED:      6-21-91
C                  7-18-91
C
C      THIS PROGRAM GENERATES A DATA FILE FOR THE FINITE ELEMENT
C      PROGRAM RBUP2.FOR
C
C
C
C
C      INTEGER SYSD,KB(24,1000),LSS,LCL,LCU,TSS,TCL,TCU,BSS,BCL,BCU
C      INTEGER RSS,RCL,RCU,BKB(8,10),MAXI(30)
C      INTEGER I,J,K,N,NE,NEM
C      REAL THETA(10)
C      WRITE(5,1)'THIS PROGRAM GENERATES A DATA INPUT FILE FOR THE'
C      WRITE(5,2)'FINITE ELEMENT CODE RBUP2.FOR'
C      WRITE(5,2)'THE DATA FILE CREATED WILL BE RBIN.DAT'
C      WRITE(5,2)'YOU WILL BE PROMPTED FOR INFORMATION'
1      FORMAT(/,4X,A)
2      FORMAT(4X,A)
3      FORMAT(4X,A,/)
C      WRITE(5,1)'ENTER THE # OF RECTANGULAR ELEMENTS THE PLATE IS TO BE'
C      WRITE(5,2)'DIVIDED INTO. THERE MUST BE AN EQUAL # OF ELEMENTS ON'
C      WRITE(5,3)'EACH SIDE. ELEMENTS HAVE 24 D.O.F.'
C      READ(5,*)NE
C      WRITE(5,1)'PLEASE ENTER THE NUMBER OF MODES TO BE INCLUDED'
C      WRITE(5,3)'IN THE NON-LINEAR FORMULATION'
C      READ(5,*)MODE
C      WRITE(5,1)'ZETA = DAMPING RATIO'
C      WRITE(5,2)'BWLFR = LOW FREQ. OF THE BAND-LIMITED WHITE NOISE (HZ)'
C      WRITE(5,2)'BWHFR = HIGH " " " " " " " " " "
C      WRITE(5,3)'ENTER ZETA,BWLFR,BWHFR (ex: 0.01,0.1,10000.0)'
C      READ(5,*)ZETA,BWLFR,BWHFR
C      WRITE(5,1)'ENTER PRESET VALUE FOR MAX # OF ITERATIONS (ex. 20)'
C      READ(5,*)ITERM
C      WRITE(5,1)'A, B = LENGTH, WIDTH OF RECTANGULAR PLATE (in.)'
C      WRITE(5,2)'RHO = MASS DENSITY (lb-sec**2/in**4)'
C      WRITE(5,2)'UM = POISSON'S RATIO'
C      WRITE(5,2)'PACN = PRE-ASSIGNED CONVERGENCE VALUE FOR SUCCESSIVE'
C      WRITE(5,2)'ITERATIONS OF RMS DISPLACEMENTS (ex: 0.001)'
C      WRITE(5,3)'ENTER: A,B,RHO,UM,PACN'
C      READ(5,*)A,B,RHO,UM,PACN
C      WRITE(5,1)'ENTER THE INITIAL AND FINAL SOUND SPECTRUM LEVELS'
C      WRITE(5,2)'AND THE INCREMENT (all in dB): INITIAL,FINAL,INCR'
C      READ(5,*)SPDBI,SPDBF,SPDBDL
C      WRITE(5,1)'IF THE PLATE IS A COMPOSITE LAMINATE, ENTER THE PLY'
C      WRITE(5,2)'MATERIAL PROPERTIES IN THE PRINCIPAL DIRECTIONS'
C      WRITE(5,2)'NOTE: IF THE PLATE IS ISOTROPIC, LET E1 = E2,'
C      WRITE(5,3)'UM12 = UM, G12 = G ec...'
C      WRITE(5,2)'ENTER: E1,E2,UM12,G12,PLATE THICKNESS,# PLIES'
C      READ(5,*)E1,E2,AMU12,G12,THICK,NLAY
C      WRITE(5,1)'START WITH TOP PLY, ENTER ORIENTATION ANGLE (deg.)'
C      READ(5,*)THETA(1)
C      J=2
C      DO WHILE(J.LE.NLAY)
C          WRITE(5,2)'ENTER ANGLE'
C          READ(5,*) THETA(J)
C          J=J+1
C      END DO

```

```

WRITE(5,1)'NOW THE PLATE BOUNDARY CONDITIONS MUST BE ENTERED'
WRITE(5,2)'THE PLATE IS DEFINED HAVING A TOP, BOTTOM, LEFT,'
WRITE(5,2)'RIGHT, WITH THE ORIGIN AT THE BOTTOM LEFT CORNER'
WRITE(5,2)'EACH SIDE MAY BE EITHER CLAMPED, SIMPLY-SUPPORTED,'
WRITE(5,2)'CUT. (CUT WOULD BE USED ONLY IF YOU ARE MODELLING'
WRITE(5,2)'A SECTION OF A LARGER PLATE)'
WRITE(5,1)'YOU NEED TO ENTER THREE NUMBERS'
WRITE(5,1)'BOTTOM EDGE      cl: 1,0,0      ss:0,1,0      cu:0,0,1'
WRITE(5,2)'ENTER THE CORRESPONDING THREE NUMBERS'
READ(5,*)BCL,BSS,BCU
WRITE(5,1)'TOP EDGE        cl: 1,0,0      ss:0,1,0      cu:0,0,1'
WRITE(5,2)'ENTER THE CORRESPONDING THREE NUMBERS'
READ(5,*)TCL,TSS,TCU
WRITE(5,1)'RIGHT EDGE     cl: 1,0,0      ss:0,1,0      cu:0,0,1'
WRITE(5,2)'ENTER THE CORRESPONDING THREE NUMBERS'
READ(5,*)RCL,RSS,RCU
WRITE(5,1)'LEFT EDGE      cl: 1,0,0      ss:0,1,0      cu:0,0,1'
WRITE(5,2)'ENTER THE CORRESPONDING THREE NUMBERS'
READ(5,*)LCL,LSS,LCU

C
C BUILT UP STRUCTURES
C
WRITE(5,1)'ARE THERE STIFFENERS ON THE PANEL ?'
WRITE(5,2)'ENTER 1 TO INCLUDE STIFFENERS.'
WRITE(5,2)'ENTER 0 TO INCLUDE NO STIFFENERS.'
READ(5,*)ISTIFF

C
C NEM IS THE NUMBER OF D.O.F. FOR THE RECT. PLATE ELEMENT
C N IS THE NUMBER OF ELEMENTS ALONG ONE SIDE OF THE PLATE
C
NEM=24
T=NE
N = NINT(SQRT(T))

C
C AL AND BL ARE THE ELEMENT DIMENSIONS
C
AL=A/N
BL=B/N

C
C IF THERE ARE NO STIFFENERS, THEN BB AND BH ARE SET TO VERY
C SMALL VALUES. THIS IS A FLAG FOR THE MAIN PROGRAM TO CARRY
C OUT THE PLATE ANALYSIS ONLY.
C
C IF THERE ARE STIFFENERS PRESENT, SUBROUTINE STIFNR IS CALLED
C TO DEFINE THE PROPERTIES AND LOCATION OF THE STIFFENER
C
IF(ISTIFF.EQ.1) THEN
  CALL STIFNR(N,AL,BL,NE,NEB,NEMB,BB,BH,E,BEL,NPLAC)
ELSE
  E=10.5E6
  BB=1.0E-9
  BH=1.0E-9
  BEL=1.0
  NEB=6
  NEMB=8
ENDIF

C
C OPEN(UNIT=6,FILE='RBIN.DAT',STATUS='NEW')
C

```

```

C
C
C   CONSTRUCT THE CONNECTIVITY MATRIX BETWEEN THE LOCAL AND SYSTEM
C   NODAL DISPLACEMENTS
C
C
C   INITIALIZE THE REGISTER VALUES. 1 INDICATES UNCONSTRAINED D.O.F.
C
      DO 11 I=1,NEM
        DO 11 J=1,NE
          KB(I,J)=1
11    CONTINUE
C
C   APPLY BOUNDARY CONDITIONS (ZERO INDICATES CONSTRAINED D.O.F.)
C
      DO 700 I=1,NE
        DO 600 K=1,N
C
C       THE CONSTRAINED DEGREES OF FREEDOM CORRESPONDING TO THE
C       GIVEN BOUNDARY CONDITIONS ARE ZEROED OUT.
C       ZEROS IN THE CONNECTIVITY MATRIX ELIMINATE
C       CONTRIBUTIONS OF CORRESPONDING ELEMENT
C       STIFFNESS TO THE SYSTEM STIFFNESS MATRIX
C
          IF(I.EQ.K) THEN
            IF(BCL.EQ.1) THEN
              DO 200 L=1,NEM,4
                KB(L,I)=0
                KB(L+1,I)=0
200          CONTINUE
              ENDIF
            IF(BSS.EQ.1) THEN
              DO 210 L=1,NEM,4
                IF(L.NE.9) THEN
                  KB(L,I)=0
                  KB(L+1,I)=0
210          CONTINUE
                ENDIF
              CONTINUE
            IF(BCU.EQ.1) THEN
              KB(9,I)=0
              KB(10,I)=0
              KB(13,I)=0
              KB(14,I)=0
              KB(21,I)=0
              KB(22,I)=0
            ENDIF
          ENDIF
          KT=N**2-N+K
          IF(I.EQ.KT) THEN
            IF(TCL.EQ.1) THEN
              DO 300 L=3,NEM,4
                KB(L,I)=0
                KB(L+1,I)=0
300          CONTINUE
            ENDIF
            IF(TSS.EQ.1) THEN
              DO 310 L=3,NEM,4
                IF(L.NE.11) THEN

```

```

        KB(L,I)=0
        KB(L+1,I)=0
        ENDIF
310      CONTINUE
      ENDIF
      IF(TCU.EQ.1) THEN
        KB(11,I)=0
        KB(12,I)=0
        KB(15,I)=0
        KB(16,I)=0
        KB(23,I)=0
        KB(24,I)=0
      ENDIF
    ENDIF
    IF(I.EQ.K*N) THEN
      IF(RCL.EQ.1) THEN
        DO 400 L=2,NEM,4
          KB(L,I)=0
          KB(L+1,I)=0
400        CONTINUE
      ENDIF
      IF(RSS.EQ.1) THEN
        DO 410 L=2,NEM,4
          IF(L.NE.6) THEN
            KB(L,I)=0
            KB(L+1,I)=0
410          ENDIF
          CONTINUE
        ENDIF
      IF(RCU.EQ.1) THEN
        KB(6,I)=0
        KB(7,I)=0
        KB(14,I)=0
        KB(15,I)=0
        KB(18,I)=0
        KB(19,I)=0
      ENDIF
    ENDIF
    IF(I.EQ.1.OR.I.EQ.K*N+1) THEN
      IF(LCL.EQ.1) THEN
        DO 500 L=1,NEM,4
          KB(L,I)=0
          KB(L+3,I)=0
500        CONTINUE
      ENDIF
      IF(LSS.EQ.1) THEN
        DO 510 L=1,NEM,4
          IF(L.NE.5) THEN
            KB(L,I)=0
            KB(L+3,I)=0
510          ENDIF
          CONTINUE
        ENDIF
      IF(LCU.EQ.1) THEN
        KB(5,I)=0
        KB(8,I)=0
        KB(13,I)=0
        KB(16,I)=0

```

```

                KB(17,I)=0
                KB(20,I)=0
            ENDIF
        ENDIF
600  CONTINUE
700  CONTINUE
C
C  ASSEMBLE ELEMENTS AND IDENTIFY SYSTEM DEGREES OF FREEDOM
C
C  THE SYSTEM DEGREES OF FREEDOM ARE LABELED BEGINNING
C  AT THE ORIGIN.  THEY ARE NUMBERED LEFT TO RIGHT FOR
C  EACH ROW OF NODES.
C  FIRST NUMBERED ARE THE W DISPLACEMENTS
C  STARTING AT THE ORIGIN AGAIN, THE SLOPES (W,x) ARE NUMBERED
C  NEXT ARE THE SLOPES (W,y)
C  IN A SIMILAR FASHION, THE TWIST (W,xy), AND THE U AND V
C  DISPLACEMENTS ARE NUMBERED
C
C
C  SYSD=1
C  DO 720 J=1,NEM,4
C  DO 710 I=1,NE
C
C      IF(KB(J,I).EQ.1) THEN
C          KB(J,I)=SYSD
C          SYSD=SYSD+1
C      ENDIF
C      DO 704 K=1,N
C          IF(I.EQ.N*K.AND.KB(J+1,I).EQ.1) THEN
C              KB(J+1,I)=SYSD
C              SYSD=SYSD+1
C          ENDIF
C      CONTINUE
C      DO 705 I=N**2-N+1,N**2
C          IF(KB(J+3,I).EQ.1) THEN
C              KB(J+3,I)=SYSD
C              SYSD=SYSD+1
C          ENDIF
C          IF(KB(J+2,I).EQ.1.AND.I.EQ.N**2) THEN
C              KB(J+2,I)=SYSD
C              SYSD=SYSD+1
C          ENDIF
C      CONTINUE
C      DO 720 CONTINUE
C
C  SYSDF IS THE TOTAL NUMBER NON-ZERO DEGREES OF FREEDOM
C
C  NS=SYSD-1
C
C  THE DEGEES OF FREEDOM AT COMMON NODES ARE EQUATED
C
C  DO 810 J=1,NEM,4
C  DO 800 I=1,NE
C  DO 790 K=1,N
C      IF(KB(J+1,I).EQ.1.AND.I.NE.N*K) THEN
C          KB(J+1,I)=KB(J,I+1)
C      ENDIF
C      IF(I.GE.N**2-N+1.AND.KB(J+2,I).EQ.1) THEN

```

```

        KB(J+2,I)=KB(J+3,I+1)
      ENDIF
790   CONTINUE
800   CONTINUE
      DO 805 I=1,N**2-N
        IF(KB(J+3,I).EQ.1) THEN
          KB(J+3,I)=KB(J,I+N)
        ENDIF
        IF(KB(J+2,I).EQ.1) THEN
          KB(J+2,I)=KB(J+1,I+N)
        ENDIF
805   CONTINUE
810   CONTINUE
C
C   LAST NUMBER OF DEFLECTION W IS THE LAST POSITION
C   IN THE NODAL DISPLACEMENT VECTOR HOLDING A
C   DISPLACEMENT VALUE. CALCULATE THE LAST NUMBERS OF
C   DISPLACEMENT (NW), SLOPE (NR), AND TWIST (NT).
C
      J=1
      DO 820 I=1,NEM,4
        MAXI(J)=MAX(KB(I,NE),KB(I+1,NE),KB(I+2,NE),KB(I+3,NE))
        J=J+1
820   CONTINUE
      NW=MAXI(1)
      NR=MAXI(3)
      NT=MAXI(4)
      ND=NW
C
C   WRITE DATA TO INPUT FILE RBIN.DAT
C
      WRITE(6,251)NE,NEM
      WRITE(6,252)MODE
      WRITE(6,249)NS,ND,NW,NR,NT
      WRITE(6,250)BCL,BSS,BCU,TCL,TSS,TCU
      WRITE(6,250)RCL,RSS,RCU,LCL,LSS,LCU
      WRITE(6,253)ZETA,BWLFR,BWHFR
      WRITE(6,252)ITERM
      WRITE(6,254)AL,BL,RHO,UM,PACN,THICK
      WRITE(6,255)SPDBI,SPDBF,SPBDL
      WRITE(6,256)BB,BH,E,BEL
      WRITE(6,251)NEMB,NEB
      WRITE(6,257)E1,E2,AMU12,G12,THICK,NLAY
C
249  FORMAT(I4,1X,I4,1X,I4,1X,I4,1X,I4)
250  FORMAT(I1,1X,I1,1X,I1,1X,I1,1X,I1,1X,I1)
251  FORMAT(I4,1X,I4)
252  FORMAT(I4)
253  FORMAT(F6.3,1X,F10.3,1X,F10.3)
254  FORMAT(F16.9,1X,F16.9,1X,E14.9,1X,F5.3,1X,F6.4,1X,F8.5)
255  FORMAT(F8.3,1X,F8.3,1X,F8.3)
256  FORMAT(F6.3,1X,F6.3,1X,E12.7,1X,F6.3)
257  FORMAT(E12.7,1X,E12.7,1X,F6.3,1X,E15.9,1X,F8.5,1X,I4)
258  FORMAT(I4)
259  FORMAT(F10.4)
      IZERO=0
      IONE=1

```

C


```

      DO 830 I=1,NLAY
        WRITE(6,259)THETA(I)
        IF(I.EQ.NLAY) THEN
          WRITE(6,258)IZERO
        ELSE
          WRITE(6,258)IONE
        ENDIF
830  CONTINUE
999  FORMAT(24I3)
      DO 721 M=1,NE
        WRITE(6,999) (KB(I,M),I=1,NEM)
721  CONTINUE
C
C  INITIALIZE CONNECTIVITY MATRIX FOR THE BEAM ELEMENTS
C
      DO 723 I=1,8
        DO 722 J=1,10
          BKB(I,J)=0
722  CONTINUE
723  CONTINUE
C
C  IF STIFFENERS ARE PRESENT, THEN SUBROUTINE BCONN WILL
C  CALCULATE THE CONNECTIVITY MATRIX FOR THE BEAM ELEMENTS
C
C  WITH NO STIFFENERS PRESENT, THE CONNECTIVITY MATRIX FOR
C  THE BEAM ELEMENTS REMAINS FILLED WITH ZEROS, AND THUS
C  BEAM ELEMENTS PROVIDE NO CONTRIBUTION TO THE SYSTEM
C  MASS AND STIFFNESS MATRICES
C
C
      IF(ISTIFF.EQ.0) GO TO 94
      CALL BCONN(BKB,KB,NPLAC,N,NE,NEMB,NEB)
94  CONTINUE
      DO 726 I=1,NEB
        WRITE(6,725) (BKB(J,I),J=1,NEMB)
726  CONTINUE
725  FORMAT(8I3)
      CLOSE(UNIT=6)
      END
C
C  *****
C  SUBROUTINE STIFNR
C  *****
C
C  SUBROUTINE STIFNR DEFINES THE STIFFENER PROPERTIES AND
C  ALLOWS THE USER TO INPUT THE LOCATION OF THE STIFFENER
C  WITH RESPECT TO THE PLATE.
C
      SUBROUTINE STIFNR(N,AL,BL,NE,NEB,NEMB,BB,BH,E,BEL,NPLAC)
        WRITE(5,101)'ENTER THE BEAM HEIGHT AND WIDTH (in.)'
        READ(5,*)BH,BB
        WRITE(5,101)'ENTER THE BEAM YOUNGS MODULUS (psi.)'
        READ(5,*)E
101  FORMAT(/,4X,A)
C
        WRITE(5,102)'THE PROGRAM IS SET UP TO PLACE STIFFENERS'
        WRITE(5,103)'PARALLEL TO THE PLATE'S Y-AXIS.'
        WRITE(5,101)'THE STIFFENER MUST BE PLACED AT ONE'

```

```

      WRITE(5,103)'OF THE EXISTING PLATE NODES ALONG'
      WRITE(5,103)'THE X-AXIS.'
102  FORMAT(///,4X,A)
103  FORMAT(4X,A)
      WRITE(5,104)'YOU HAVE CHOSEN TO DIVIDE THE PLATE INTO',NE
      WRITE(5,105)'ELEMENTS. WITH THIS CONFIGURATION, THERE'
      WRITE(5,105)'ARE NODES AT THE FOLLOWING LOCATIONS ON'
      WRITE(5,105)'THE X-AXIS:'
104  FORMAT(/,4X,A,1X,I3)
105  FORMAT(4X,A)
C
      XLOC=AL
      DO 10 I=1,N-1
          WRITE(5,110)I,XLOC
          XLOC=XLOC+AL
10   CONTINUE
      I = I-1
      WRITE(5,115)'ENTER THE CORRESPONDING # TO PLACE THE'
      WRITE(5,117)'STIFFENER. (ie. 1 -',I
      READ(5,*)NPLAC
C
C  NEMB IS THE NUMBER OF D.O.F. OF THE BEAM ELEMENT
C  NEB IS THE NUMBER OF BEAM ELEMENTS
C  BEL IS THE LENGTH OF EACH BEAM ELEMENT. DUE TO THE CONFIGURATION
C  OF THE STIFFENERS, BEL IS EQUAL TO THE LENGTH (BL) OF EACH PLATE
C  ELEMENT.
C
      NEMB = 8
      NEB = N
      BEL = BL
110  FORMAT(10X,I,':',4X,'X =',1X,F6.2)
115  FORMAT(/,4X,A)
117  FORMAT(4X,A,1X,I2,')')
      RETURN
      END
C
C *****
C      SUBROUTINE BCONN
C *****
C
C  THIS SUBROUTINE TAKES AS INPUT THE PLACEMENT POSITION
C  (NPLAC) OF THE STIFFENER AND CREATES THE CONNECTIVITY
C  MATRIX BETWEEN THE LOCAL BEAM DEGREES OF FREEDOM AND
C  THE SYSTEM DEGREES OF FREEDOM. THE STIFFENER AND PLATE
C  ARE ASSEMBLED TOGETHER BY EQUATING THE SYSTEM DEGREES
C  OF FREEDOM OF THE TWO AT EACH COMMON NODE. THE LOCAL
C  DEGREES OF FREEDOM CORRESPOND AS FOLLOWS:
C
C          BEAM      PLATE
C          ----      -
C          W          W
C          Wx         = Wy
C          phi        = -Wx
C          U          = V
C
C
C
      SUBROUTINE BCONN(BKB,KB,NPLAC,N,NE,NEMB,NEB)
      INTEGER BKB(NEMB.NEB),KB(24,NE)

```

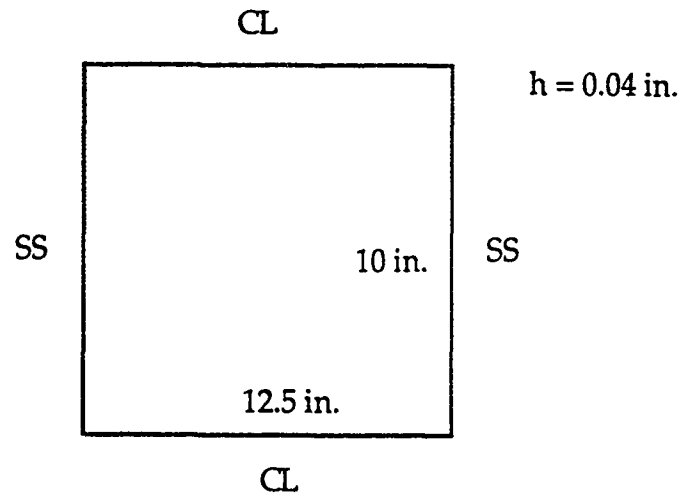
```

      NUM = NPLAC
      DO 100 I=1,NEB
        BKB(1,I)=KB(22,NUM)
        BKB(2,I)=KB(23,NUM)
        BKB(3,I)=KB(2,NUM)
        BKB(4,I)=KB(10,NUM)
        BKB(5,I)=KB(3,NUM)
        BKB(6,I)=KB(11,NUM)
        BKB(7,I)=KB(6,NUM)
        BKB(8,I)=KB(7,NUM)
      NUM = NPLAC + N*I
100   CONTINUE
      RETURN
      END

```

APPENDIX B:
SAMPLE INPUT DATA FILE

PROPERTIES USED FOR SAMPLE INPUT FILE



Aluminum Plate

density = $.2588 \times 10^{-3}$ lb sec/in⁴

Poisson's ratio = 0.33

Young's modulus = 10.5×10^6 psi

```

36 24
6
160 25 25 85 110
1 0 0 1 0 0
0 1 0 0 1 0
0.010 0.100 10000.000
10
2.500000000 2.000000000 .258799992E-03 0.330 0.0010 0.04000
90.000 90.000 1.000
0.000 0.000 .1050000E+08 1.000
8 6
.1050000E+08 .1050000E+08 0.330 0.394700000E+07 0.04000 3
0.0000
1
0.0000
1
0.0000
0
0 0 1 0 0 0 27 26 0 0 61 0 0 0 86 0 0 0111 0 0 0136 0
0 0 2 1 0 0 28 27 0 0 62 61 0 0 87 86 0 0112111 0 0137136
0 0 3 2 0 0 29 28 0 0 63 62 0 0 88 87 0 0113112 0 0138137
0 0 4 3 0 0 30 29 0 0 64 63 0 0 89 88 0 0114113 0 0139138
0 0 5 4 0 0 31 30 0 0 65 64 0 0 90 89 0 0115114 0 0140139
0 0 0 5 0 0 32 31 0 0 0 65 0 0 0 90 0 0 0115 0 0 0140
0 1 6 0 26 27 34 33 0 61 66 0 0 86 91 0 0111116 0 0136141 0
1 2 7 6 27 28 35 34 61 62 67 66 86 87 92 91111112117116136137142141
2 3 8 7 28 29 36 35 62 63 68 67 87 88 93 92112113118117137138143142
3 4 9 8 29 30 37 36 63 64 69 68 88 89 94 93113114119118138139144143
4 5 10 9 30 31 38 37 64 65 70 69 89 90 95 94114115120119139140145144
5 0 0 10 31 32 39 38 65 0 0 70 90 0 0 95115 0 0120140 0 0145
0 6 11 0 33 34 41 40 0 66 71 0 0 91 96 0 0116121 0 0141146 0
6 7 12 11 34 35 42 41 66 67 72 71 91 92 97 96116117122121141142147146
7 8 13 12 35 36 43 42 67 68 73 72 92 93 98 97117118123122142143148147
8 9 14 13 36 37 44 43 68 69 74 73 93 94 99 98118119124123143144149148
9 10 15 14 37 38 45 44 69 70 75 74 94 95100 99119120125124144145150149
10 0 0 15 38 39 46 45 70 0 0 75 95 0 0100120 0 0125145 0 0150
0 11 16 0 40 41 48 47 0 71 76 0 0 96101 0 0121126 0 0146151 0
11 12 17 16 41 42 49 48 71 72 77 76 96 97102101121122127126146147152151
12 13 18 17 42 43 50 49 72 73 78 77 97 98103102122123128127147148153152
13 14 19 18 43 44 51 50 73 74 79 78 98 99104103123124129128148149154153
14 15 20 19 44 45 52 51 74 75 80 79 99100105104124125130129149150155154
15 0 0 20 45 46 53 52 75 0 0 80100 0 0105125 0 0130150 0 0155
0 16 21 0 47 48 55 54 0 76 81 0 0101106 0 0126131 0 0151156 0
16 17 22 21 48 49 56 55 76 77 82 81101102107106126127132131151152157156
17 18 23 22 49 50 57 56 77 78 83 82102103108107127128133132152153158157
18 19 24 23 50 51 58 57 78 79 84 83103104109108128129134133153154159158
19 20 25 24 51 52 59 58 79 80 85 84104105110109129130135134154155160159
20 0 0 25 52 53 60 59 80 0 0 85105 0 0110130 0 0135155 0 0160
0 21 0 0 54 55 0 0 0 81 0 0 0106 0 0 0131 0 0 0156 0 0
21 22 0 0 55 56 0 0 81 82 0 0106107 0 0131132 0 0156157 0 0
22 23 0 0 56 57 0 0 82 83 0 0107108 0 0132133 0 0157158 0 0
23 24 0 0 57 58 0 0 83 84 0 0108109 0 0133134 0 0158159 0 0
24 25 0 0 58 59 0 0 84 85 0 0109110 0 0134135 0 0159160 0 0
25 0 0 0 59 60 0 0 85 0 0 0110 0 0 0135 0 0 0160 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0

```

0 0 0 0 0 0 0 0

APPENDIX C:
SAMPLE OUTPUT DATA FILE

On File At RDL

APPENDIX D:
PROGRAM LISTING: MODPLT.PRO
MODE SHAPE PLOTS

```

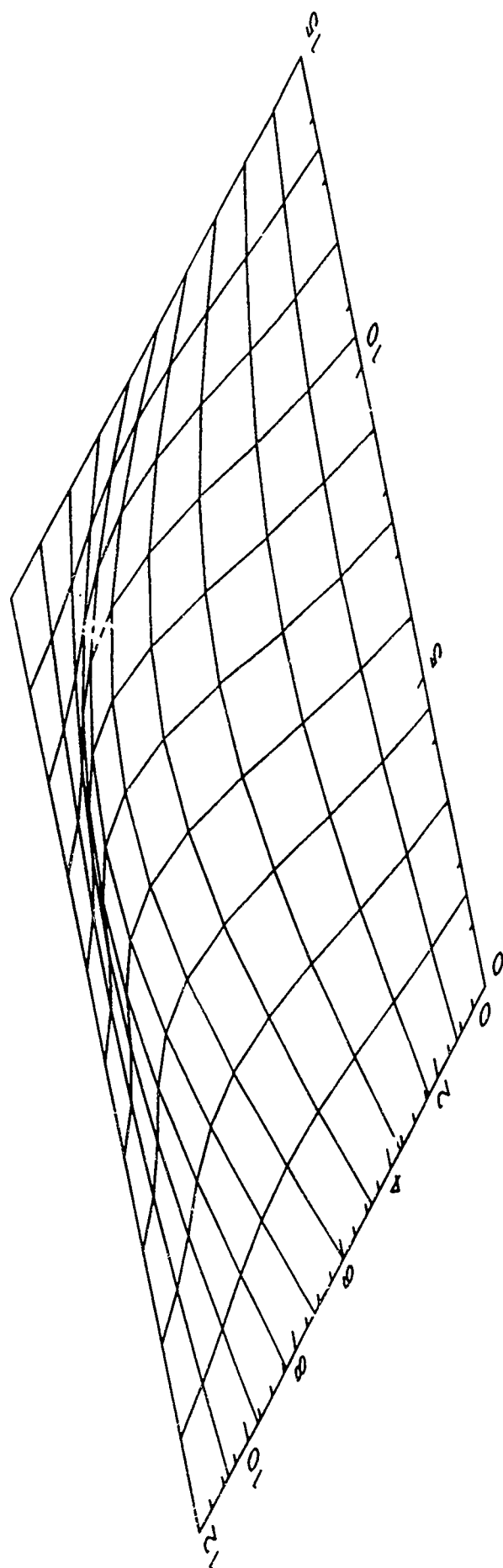
;
; PROGRAM:          MODPLT.PRO
; WRITTEN BY:       RUSSELL A. MORRIS
; DATE:            7-31-91
;
; THIS FILE IS TO BE RUN WHILE INSIDE PV-WAVE
; AT THE WAVE PROMPT, TYPE: .RUN MODPLT
; THIS PROGRAM PLOTS MODE SHAPES FROM THE
; FILE MODE.DAT
      OPENR,1,'MODE.DAT'
      READF,1,N
      X=FLTARR(N)
      Y=FLTARR(N)
      Z=FLTARR(N,N)
;
; NOW READ IN DATA
      READF,1,X
      READF,1,Y
      READF,1,Z
;
; POSITION PLOT ON PAGE
      !P.POSITION=[0.08,0.08,.95,.95]
      !P.CHARSIZE=1.4
      !P.REGION=[0.1,0.39,0.89,0.81]
;
; DRAW SURFACE PLOT, FIRST CREATE PRINTER READY VERSION
      !P.COLOR=5
      LN03,'MODPLT.LN3'
      SURFACE,Z,X,Y,ZAX=-1,ZRAN=[0,5],AX=30
      XYOUTS,500,2150,'ALUMINUM PLATE',/DEVICE
      XYOUTS,500,2050,'SS-CL-SS-CL',/DEVICE
      XYOUTS,500,1750,'FREQUENCY (rad/sec) =',/DEVICE
      XYOUTS,1400,2600,'LINEAR MODE SHAPE',/DEVICE,SIZE=1.6
      LN03
;
; TO PRINT HARD COPY, FROM WITHIN WAVE:
; TYPE: $PRINT/PASSALL/QUE=TTA3: MODPLT.LN3
;
; NOW DRAW TO SCREEN
      SURFACE,Z,X,Y,ZAX=-1,ZRAN=[0,5],AX=30
      XYOUTS,90,300,'ALUMINUM PLATE',/DEVICE
      XYOUTS,90,285,'SS-CL-SS-CL',/DEVICE
      XYOUTS,90,240,'FREQUENCY (rad/sec) =',/DEVICE
      XYOUTS,300,450,'LINEAR MODE SHAPE',/DEVICE,SIZE=1.6
      CLOSE,1
      END

```

LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

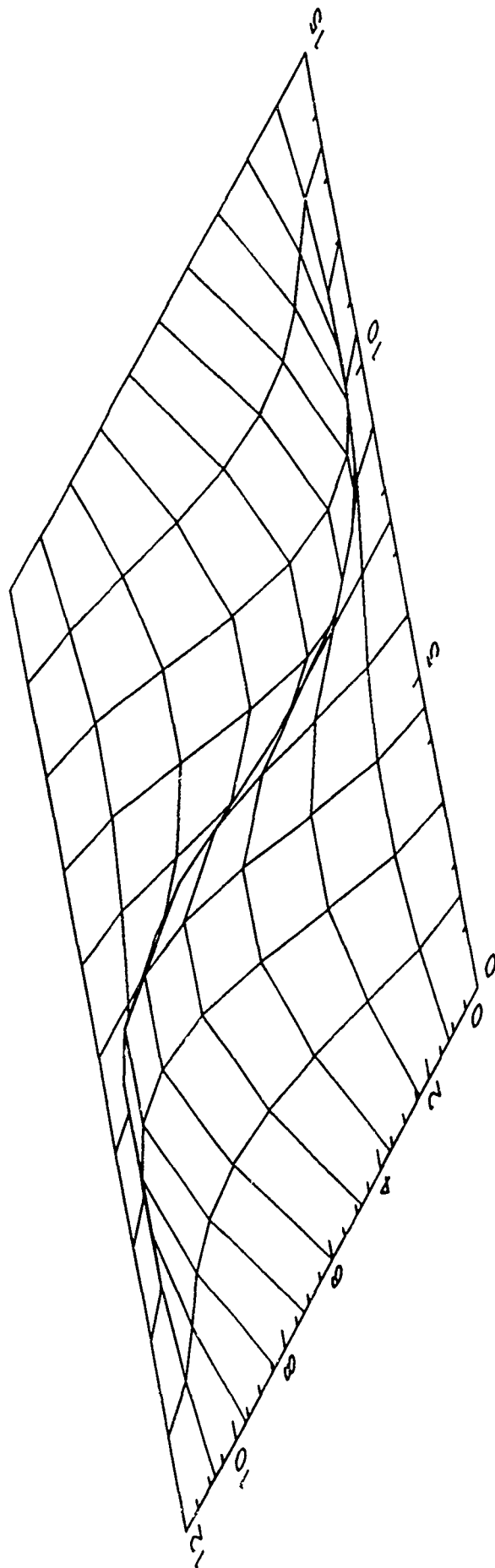
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

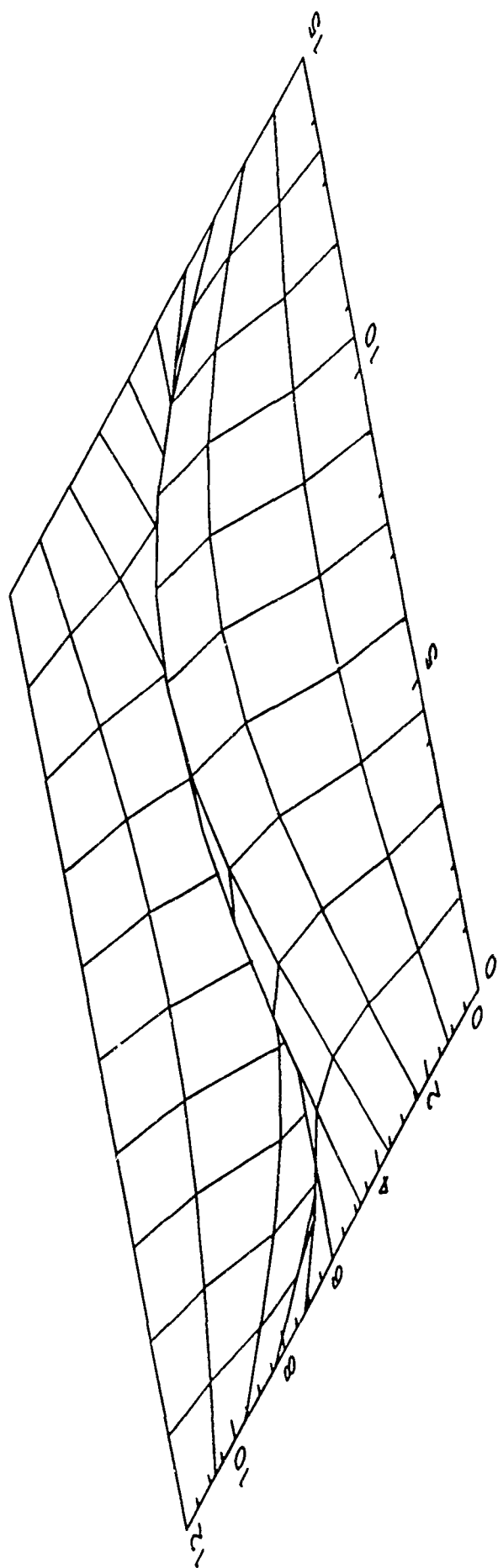
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

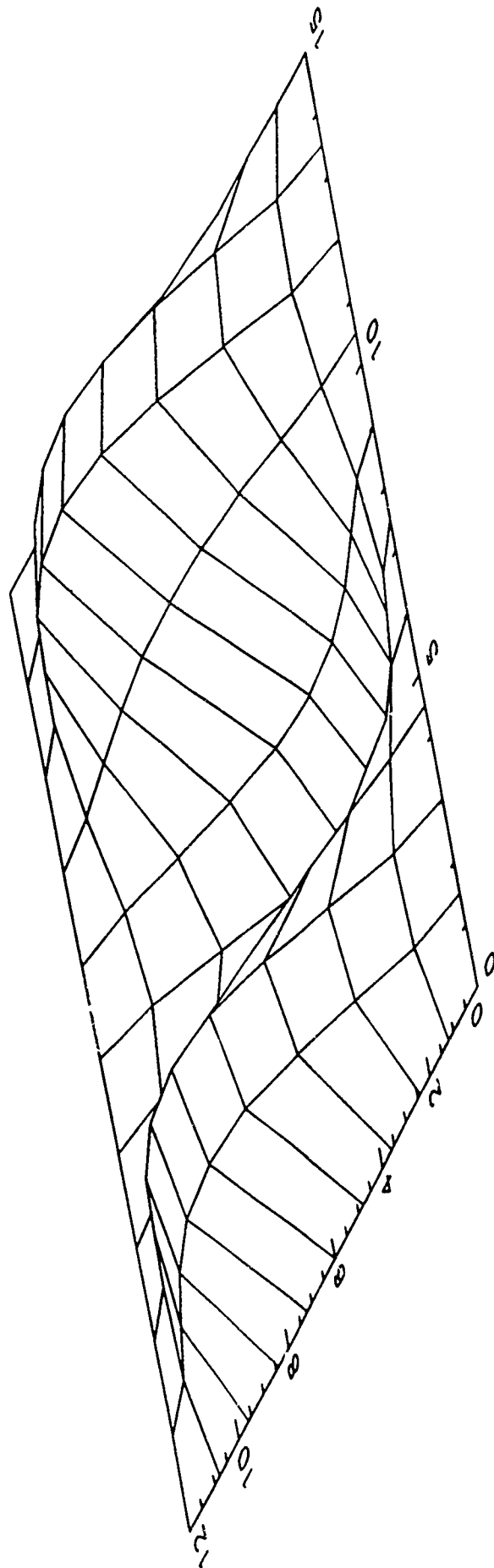
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

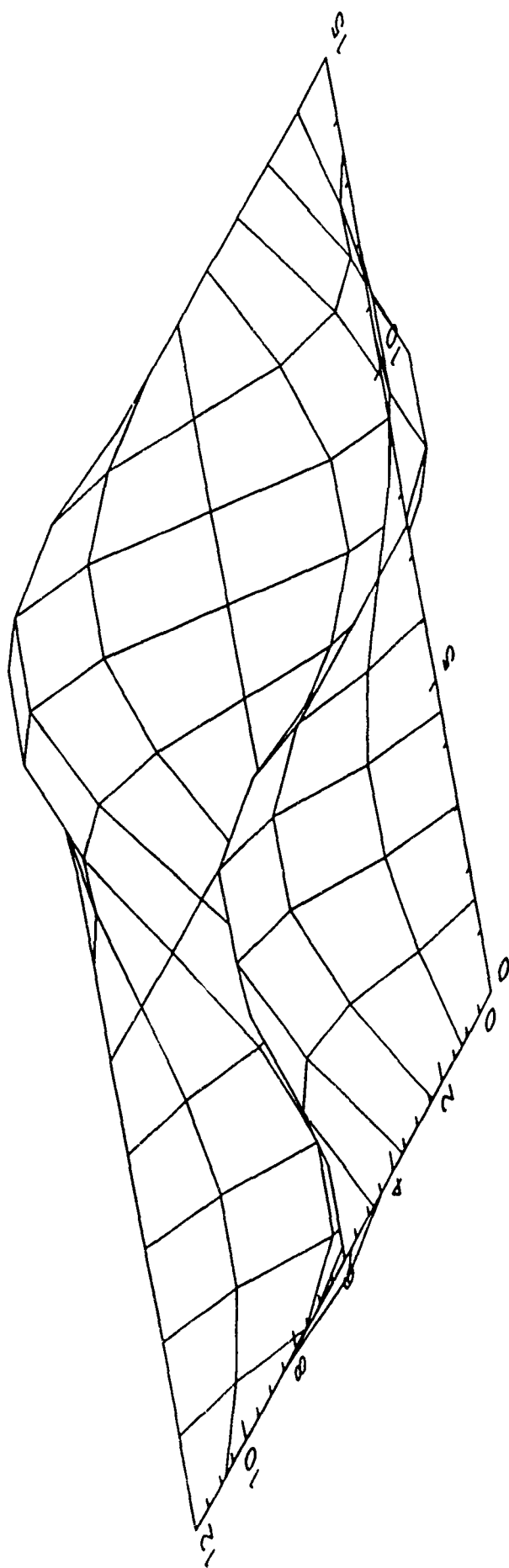
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

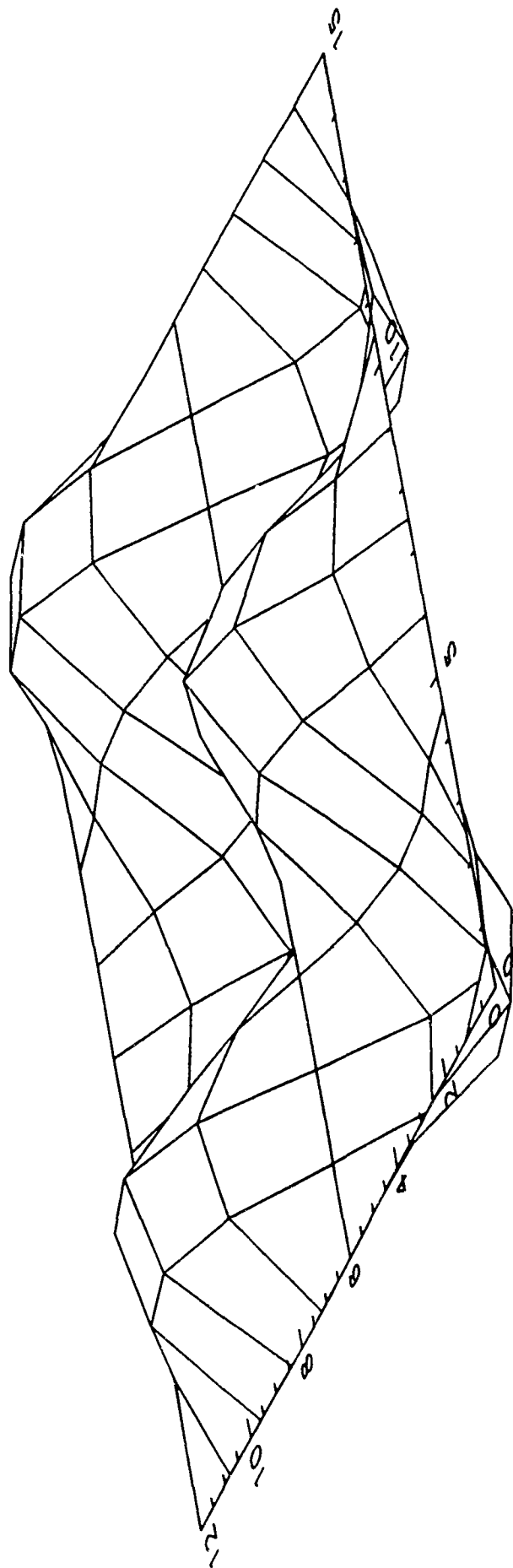
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

ALUMINUM PLATE
SS-CL-SS-CL

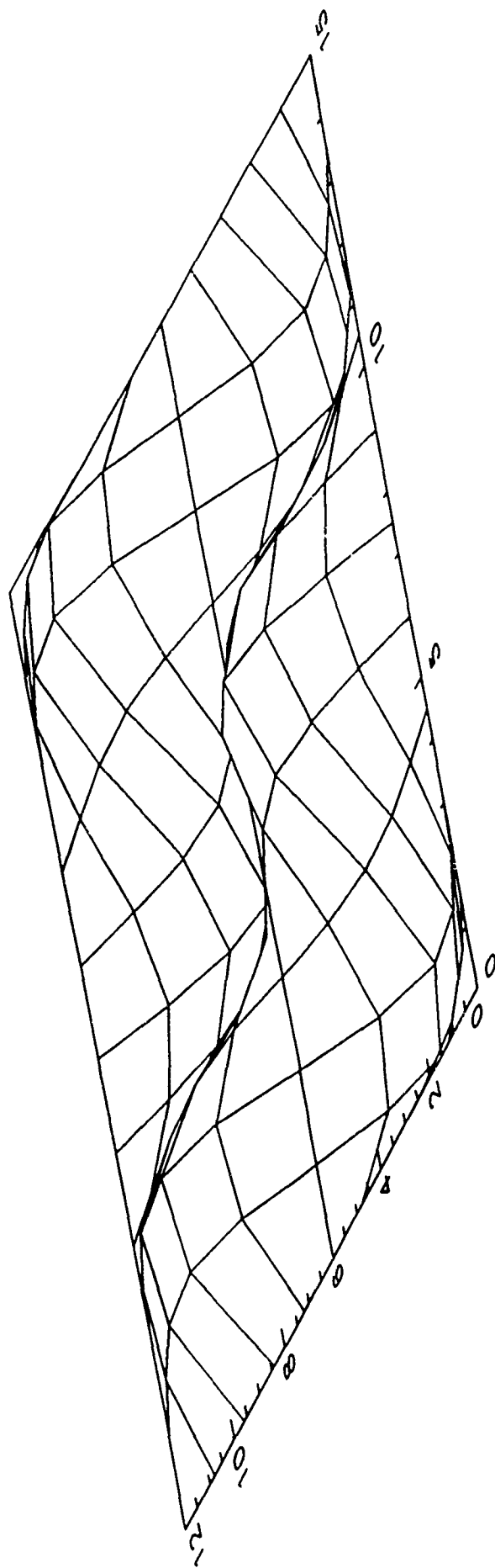
FREQUENCY (rad/sec) =



LINEAR MODE SHAPE

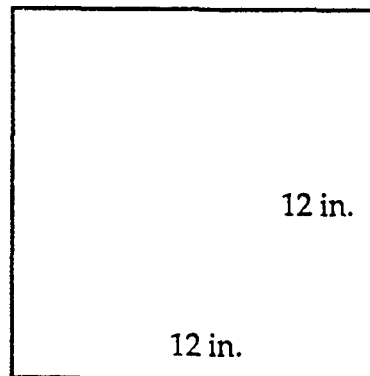
ALUMINUM PLATE
SS-CL-SS-CL

FREQUENCY (rad/sec) =



APPENDIX E:
CONVERGENCE STUDY MATERIAL PROPERTIES

PLATE DIMENSIONS AND PROPERTIES FOR CONVERGENCE STUDY



$h = 0.04$ in. (thickness)

ALUMINUM PLATE

density = $.2588e-3$ lb sec²/in⁴

Poisson's ratio = 0.33

Young's modulus = $10.5e-6$ psi

GRAPHITE-EPOXY PLATE

$$\begin{bmatrix} 0 \\ -45 \\ 45 \\ 90 \end{bmatrix}_s$$

density = $.1509e-3$ lb sec²/in⁴

$E_1 = 20.46e6$ psi

$E_2 = 1.34e6$ psi

$G_{12} = 8.638e5$ psi

$\nu_{12} = 0.3131$

EVALUATION OF THE ANALYTICAL DESIGN PACKAGE (ADP)
FOR FRAMELESS TRANSPARENCY PROGRAM

David M. Russell, Ph.D. Student
Joe Chow, Assistant Professor
Clarkson University, Potsdam, New York

Acknowledgment

We are very grateful to Group Leader Robert McCarty, Program Leader William Pinnel, and Aerospace Engineer Kevin Roach for their support and for providing us with a comfortable working atmosphere. In particular, we would like to thank Kevin for his help and guidance throughout this research period and efforts in communicating our findings to General Dynamics. The Flight Dynamics Laboratory and the Air Force Office of Scientific Research are also appreciated for the sponsorship of this research. Finally, we want to thank the people of the Aircrew Protection Branch, who have made this summer a rewarding and fun experience.

Abstract

The increasing demand for more efficient transparency design, low cost manufacturing, and longer service life has lead to the concept of frameless transparencies for high performance fighter and trainer type aircraft. The frameless transparency program investigates the use of a transparent panel which directly interfaces with the aircraft structure without using a frame. Computer programs, such as Patran, MAGNA, STAPAT, and C-Mold, have been used to create, assess and optimize traditional transparency designs. However, these stand alone programs can be difficult to run, as the user is expected to be quite proficient with all procedures for running these programs. To aid in the transparency design procedure, an Analytical Design Package, ADP, is being developed by General Dynamics (GD), which integrates these computer programs into one module. The main

objective of this project is to test and validate the ADP, and ensure that it is easier to use compared to using the stand alone analysis programs. This involves the evaluation of the ADP's user interface and translation programs. The procedures used in evaluating the ADP and the results of the evaluation are documented in this report.

1. Introduction

Presently the windows and canopies for high performance fighter and trainer type aircraft are made of transparency assemblies which, consist of transparent panels mounted in reinforcing metal frames. The frame serves as an interface between the transparent panel and the aircraft structure. The manufacture of these transparency assemblies involves heating previously extruded flat sheets and forcing them to conform to a mandrel, or tool, which has the desired contour. After going through the proper heating cycles, the desired shape is achieved and retained. The transparent panel is then edge trimmed and mounted in a frame. The presence of the frame reduces the structural integrity and dramatically increases the cost of manufacturing and maintenance. The manufacture and implementation of a frameless transparency program would result in projected savings of over \$600M for the F-16 fleet alone. But, the cost of molds for frameless transparencies requires that the designers fully investigate a design before a mold is made.

New injection molding processes have demonstrated potential for directly forming transparencies from plastic resins with attachment hardware directly integrated into a molded transparency [1]. These new processes form frameless transparencies by injecting molten resins into a high precision mold. This fabrication method allows for high precision thickness control, achieving superior optical quality in viewing areas, edge reinforcement, and accommodation of transparency to aircraft

interface. In order to develop and validate this process, a frameless transparency program has been initiated by the Aircrew Enclosures Group at Wright Laboratory (WL). An integral part of this project is the ADP, which was developed to aid in producing a frameless transparency design.

The ADP is a design tool which integrates the necessary analysis tools, used to iteratively converge on a transparency design. It is capable of producing a frameless transparency design with integrated aircraft interface, mold design, and molding process parameters. The four major areas of the transparency design process are unified by the ADP, these areas are; computer aided design (CAD), structural finite element analysis, aerodynamic heating and transient thermal analysis, and injection molding analysis. The programs used in these areas are, Patran, MAGNA (Materially and Geometrically Nonlinear Analysis), STAPAT (Specific Thermal Analyzer Program for Aircraft Transparencies), and C-Mold, respectively. The integration of the different analysis programs allows the user to converge on a design quickly, reducing the time and cost required to produce a frameless transparency design.

2. Research Objective

The objective of this research is to test and evaluate the ADP. This package, designed and written by General Dynamics (GD), integrates the analysis programs used by the Aircrew Enclosure Group at Wright Laboratory, (WL/FIVR). The analysis programs incorporated in the ADP have been validated by the original software developers and/or WL. Therefore, the emphasis of this project is to evaluate the ADP's user interface, translation programs, and the general overall operation of the package. The final goal of this project is to ensure that the ADP is user friendly, operates smoothly, and correctly shares data between the various analysis programs.

3. ADP Evaluation Procedure

The evaluation of the package was divided into three major phases;

1. Test the menu system and user interface,
2. Check the data translation programs, and
3. Run a transparency through a complete design cycle in the ADP.

The activities during each of these phases are described in the following subsections of this report. The ADP is still in the development stage and had to be assessed accordingly. This required an approach which evaluated and tested the high level functionality, (user interface), before the low level translating programs. The idea being that, it doesn't matter if the data can be translated between programs if the user can't run the programs. As errors were encountered, Software Incident Reports, SIRs, were created and passed on to GD, where the problems could be fixed.

The core program of the ADP is the Executive Control Processor (ECP) program. The ECP is the root menuing system which consolidates the four main computer programs, (Patran, MAGNA, STAPAT, and C-Mold), the file translators, and a materials database manager, (FTPMAT). The ECP organizes related options, according to these tasks, into a tree structure menu system, and, can be run using a textual or graphical mode. The graphical mode of operation was used exclusively through this evaluation procedure.

3.1 Testing the ADP's Menu System

The first goal in evaluating the ADP was to make sure that the menu system is easy to use, well organized, and self explanatory. Most of this evaluation was performed on ADP Version 1.1, (ADP V1.1), but continued through the later versions.

Evaluation of the ADP V1.1 began (1 July 91), by traversing all possible menu paths, without running any analysis. Any difficulties and errors encountered were documented. These errors were organized into a report which was sent on to GD (11 July 91). This report resulted in 48 SIRs, of which 13 still remain unresolved (Sept 91). As later versions of the ADP were delivered, all the SIRs which remained open were checked to see if they had been resolved. Due to the vast extent of problems with the ECP portion of the ADP, it was hard to proceed with evaluating the ADP further.

When ADP V1.2 was delivered (8 July 91), it was evaluated in a manner similar to ADP V1.1. This was done to see which problems had been resolved and locate any new errors. This evaluation resulted in the creation of 11 SIRs (17 July 91), of which 4 still remain unresolved (Sept 91). The delivery of ADP V1.2a (19 July 91), was a vast improvement over the previous versions and allowed for greater testing of the ADP's functionality, e.g. data translations.

3.2 Data Translation Verification

The main objective of this evaluation phase was to test the ADP's data translation programs. One of GD's major efforts, in the development of the ADP, is the development and/or modification of file translators between the various analysis programs. The ADP uses Patran for the and post processing of the analysis models. By running various analysis programs on sample models, created in Patran, one can find out whether or not the file translators between various analysis programs are operating correctly.

Two models were used for testing the ADP's ability to set up an analysis, run the analysis, and then review the results. One of the models used for testing at this level was a confirmation frameless transparency,

CFT. The other model was a much simpler model, DMR. This model was a crude estimation of the CFT. These models were used to test the various data translation tasks within the ADP because, the analysis would run quicker and the cause of any problems would be easier to locate. In order to test all of the translating programs, it was necessary to perform the analyses and subsequent post processing of the results. The validity of the results obtained was not as important as demonstrating that the ADP was able to process the data. The general procedure for running MAGNA, STAPAT (STAHET and TAP), and C-Mold programs inside the ADP is as follows;

- (1) On Iris workstation prepare the analysis model inside Patran.
- (2) Prepare input files for the appropriate analysis programs.
- (3) Run the analysis.
- (4) Convert the analysis results to a Patran format.
- (5) Review results inside Patran.

In the case of running MAGNA and STAPAT one had to also be concerned with the data transfer between the Iris Workstation and the Cray computer, where the MAGNA and STAPAT programs reside.

Evaluation of the ADP translation programs began with version 1.2, and continued with version 1.2a. The unresolved short comings of the ADP V1.1 made evaluation at this level difficult. Even with the problems that still existed in the ECP portion of the ADP, it became obvious that GD had put much of their effort into the data translation between the different analysis codes. After the arrival of ADP V1.2a only 12 new SIRs were written (16 Aug 91) of which only 3 remain unresolved (Sep 91).

The evaluation of the ADP up through version 1.2a resulted in a total of 71 SIRs (16 Aug 91). Of these SIRs, 20 still have to be resolved. Of these unresolved SIRs, 7 are considered to be serious. There has been much

The design process began with creating the forebody mesh for STAHET and a canopy mesh suitable for TAP. STAHET and TAP were then run and the results viewed in Patran. (Figures depicting various results are at the end of this report.) Figure 1 shows the STAHET results in the area of the forebody where the canopy resides, and Figure 2 shows the TAP results at the end of a simple flight mission. (The left side of Figure 2 shows the canopy with the front portion cut away.) STAHET and TAP were run completely through the ADP, with relatively little difficulty. These results were then applied to a MAGNA mesh of the same canopy.

MAGNA requires a different mesh than TAP, so a new mesh had to be created. After creating a new mesh for MAGNA, simple boundary conditions were applied. The canopy centerline was restrained to model the symmetry and the edge was clamped. An appropriate loading condition was then imposed. (6 psi was applied to the interior surface, to simulate cabin pressure.) The STAHET pressures and the TAP temperatures were then applied to the model using the PCL programs STAHET_MAGNA and TAP_MAGNA, respectively. This is when the most trouble was encountered, during this level of evaluation of the ADP. Approximately 6 new serious SIRs were created as a result of these PCL programs. After quite a bit of side stepping and a few tricks, (to resolve problems with the PCL), the STAHET and TAP results were successfully applied to the MAGNA mesh. At this point static analysis of the MAGNA mesh was performed. Figure 3 shows the shaded contours of the deflections as a result of this static analysis.

With the static analysis performed, the next task was to perform a birdstrike analysis. (MAGNA has the capability of modeling the highly nonlinear effects of a bird striking a canopy at flight velocities.) Before performing this analysis, the proper time integration step size had to be determined. This was accomplished by running a frequency analysis in MAGNA. (The second mode shape from this frequency analysis is shown in

Figure 4.) With this frequency analysis performed, the proper step size can be determined and used for the birdstrike analysis. This analysis was then performed, and the results, (von Mises stresses on the front portion of the deformed model), at 0.001 seconds into the birdstrike are displayed in Figure 5. Other than the problems encountered in translating the TAP and STAHET results into the MAGNA mesh, the only major problem with running MAGNA within the ADP is the omission of the material properties from the MAGNA input deck. For this analysis, the material properties were input manually.

With the MAGNA analysis complete the only thing left to test was C-Mold. The testing of C-Mold was the easiest part of the ADP's evaluation. Other than creating the model in Patran, everything is done entirely within C-Mold, including material selection. C-Mold however, requires a completely different mesh made entirely of shell or plate elements. Evaluation of C-Mold during this phase was limited to the translation of the mesh, created in Patran, into C-Mold.

Testing the ADP during this phase of the evaluation resulted in the creation of 24 new SIRs (Sept 91). Of these SIRs, 5 are related to the PCL and 5 to the materials database. Of these 24 SIRs, 18 are considered to be serious.

4. Concluding Remarks:

The evaluation of the ADP's user interface, translating programs, and overall functionality has resulted in a total of 95 SIRs. Of these 95 SIRs, 44 remain to be resolved (Sept 91). 25 of these unresolved SIRs are considered to be serious. The materials database accounts for the largest part of these open SIRs, with 10 to be resolved, 6 of which are serious.

The PCL programs, e.g. STAHET_MAGNA and TAP_MAGNA, account for 7 more of the unresolved serious SIRs. The appendix of this report contains a summary/status document, which was created to aid in tracking the SIRs.

During the period of this evaluation (1 July through 6 Sept 91), the usability of the ADP has been greatly improved. The status of the ADP usability is summarized by the following statements;

1. The ECP is useable and nearly complete.
2. STAHET and TAP can be run with relatively little difficulty.
(Most of the problems should be easily fixed.)
3. Only minor problems exist in translating a Patran mesh into C-Mold. (C-Mold functionality within the ADP was not further tested.)
4. The PCL needs to be more completely tested. It is a source of many problems and confusion.
5. MAGNA can not be run using the ADP. The user must find ways to resolve the serious problems with the PCL and materials database.
6. The materials database manager FTPMAT is the greatest source of problems. (This area needs to be completely investigated in order to come up with an alternative solution.)
7. When the unresolved SIRs are fixed, development of the ADP should be virtually complete.

Compared to stand alone analysis programs, a fully operational ADP has the following advantages;

- 1) It is an integrated package, thus allowing the user to move from one analysis program to another easily.
- 2) The ADP is menu driven and most of the options are fairly self explanatory.
- 3) The procedure for running an analysis is significantly simplified. Most of the steps, which are normally done manually,

have been automatd by the ADP.

- 4) In general, the user needs to know less about the individual analysis programs, in order to run them.

Although currently there are many problems with the ADP it has shown a lot of potential for becoming an excellent design tool, capable of producing a frameless transparency design. Patran, MAGNA, STAHET and TAP, and C-Mold are successfully unified by the ADP. Integration of the different analysis programs allows the user to converge on a design quickly. And, a designer could be quickly and easily trained to use the ADP when the ADP becomes fully functional. The general feeling about the ADP is; When the remaining SIRs and the materials database problems are resolved, WL/FIVR will have a design tool which will greatly aid in the design process of a frameless transparency.

References:

1. Fuller, D. B., "Direct Forming Processes for Aircraft Transparencies", Loral Defense Systems-Arizona, Conference on Aerospace Transparency Materials and Enclosures Vol. 1, April 1989, pp 522-554, WRDC-TR-89-4044.

Appendix:

WL/FIVR SIR STATUS SUMMARY

Legend --- S# -- Suggested changes or non-critical errors.
C# -- Serious errors which must be corrected.
E# -- CRITICAL errors.
FIVR SIR# -- FIVR's SIR tracking number.
WL# -- GD's # for tracking WL originated SIRs.

S3 ---- Closed. WL1.1.27, WL1.1.29
S4 ---- Closed. WL1.1.3
S5 ---- Closed. WL1.1.5
S6 ---- Closed. WL1.1.10
S7 ---- Closed. WL1.1.29
S8 ---- Closed. WL1.1.32, WL1.1.39
S9 ---- Closed.
S11 --- Closed.
S12 --- Closed. WL1.1.41
S13 --- Closed. WL1.1.44
S14 --- Closed. WL1.1.45
S16 --- Closed.
S17 --- Closed. WL1.1.51
S18 --- Closed. WL1.1.53
S19 --- Closed. WL1.1.53
S20 --- Closed. WL1.1.55
C1 ---- Closed. WL1.1.6, WL1.1.36
C2 ---- Closed. WL1.1.9, WL1.1.21, WL1.1.38
C3 ---- Closed. WL1.1.28
C4 ---- Closed. WL1.1.14
C5 ---- Closed. WL1.1.20
C6 ---- Closed. WL1.1.25
C7 ---- Closed.
C9 ---- Closed. WL1.1.33
E1 ---- Closed. WL1.1.66
E2 ---- Closed. WL1.1.67
E3 ---- Closed. WL1.1.26
E4 ---- Closed. WL1.1.68
E5 ---- Closed. WL1.1.29
E6 ---- Closed. WL1.1.30
E7 ---- Closed. WL1.1.34
E8 ---- Closed. WL1.1.37
E9 ---- Closed. WL1.1.42
E10 --- Closed. FIVR SIR16, WL1.2.78
E11 --- Closed. WL1.1.54

S1 ---- OPEN. Use of '...'.
FIVR SIR17, WL1.2.77, 1.1.1
S2 ---- OPEN. Use of '...'.
FIVR SIR17, WL1.2.77, WL1.1.2, WL1.1.4, WL1.1.11,
WL1.1.31, WL1.1.35, WL1.1.48

S10 --- OPEN. A centralized default database is to be established.
FIVR SIR18, WL1.2.63
S15 --- OPEN. FTPMAT Browse material, menu order.
WL1.1.46
S24 --- OPEN. Each window should have its own title.
FIVR SIR15, WL1.2.80
S25 --- OPEN. PATMAG should reference MAGNA variable names.
WL1.1.71
S26 --- OPEN. Work directory name to be put on the second line of the
status/input box.

FIVR SIR4, WL1.2.84

S27 --- OPEN. NEXT and PREV window debate settled through the reorganization of PATSTAHET data input. Specifically submenus for Mission Data.
FIVR SIR10, WL1.1.8, WL1.1.17, WL1.1.43, WL1.2.79

C8 ---- OPEN. Cancel instability in FTPMAT.
WL1.1.40

C10 --- OPEN. Some kind of filter is needed. Even backspaces are treated as separate characters.
FIVR SIR19, WL1.1.12, WL1.1.22, WL1.2.76, WL1.2.99

C11 --- OPEN. FTPMAT filename defaults use #'s.
WL1.1.47

C12 --- OPEN. List of files is wrong.
FIVR SIR9, WL1.2.81

C13 --- OPEN. Cancel doesn't work. File 'c' is used, etc..
FIVR SIR8, WL1.2.82

----- New Errors V1.2 -----

E12 --- Closed. WL1.2.49

E13 --- Closed. FIVR SIR1, WL1.2.72

E14 --- Closed. FIVR SIR2, WL1.2.71

E15 --- Closed. FIVR SIR5, WL1.2.85

E16 --- Closed. FIVR SIR12, WL1.2.65

E17 --- Closed. FIVR SIR13, WL1.2.70

E18 --- Closed. FIVR SIR14, WL1.2.69

S21 --- OPEN. Input of time for mission, in PATTAP, it is unclear.
FIVR SIR6, WL1.2.66

S28 --- OPEN. PATCMOLD and PATCCOOL are the only two places the option of inputting the Neutral file name is not given.
FIVR SIR3, WL1.2.87

S29 --- OPEN. Make changing directories permanent.
FIVR SIR7, WL1.2.83

E19 --- OPEN. FTPMAT Memory Fault coredump.
FIVR SIR11, WL1.2.64

----- New Errors V1.2a -----

S22 --- Closed. FIVR SIR30, WL1.2.96

S23 --- Open/Closed. Source code. FIVR SIR22, WL1.2.62

E20 --- Closed. FIVR SIR21, WL1.2.88

E21 --- Closed. FIVR SIR26, WL1.2.93,

E22 --- Closed. FIVR SIR27, WL1.2.90, WL1.2.90

E23 --- Closed. FIVR SIR29, WL1.2.95

E24 --- Closed. FIVR SIR31, WL1.2.97

E25 --- Closed. FIVR SIR32, WL1.2.98

E26 --- Closed. FIVR SIR23

S30 --- OPEN. In PATTAP there is a message window that 'flys' by and can't be read. To be mentioned in the manual.
FIVR SIR24, WL1.2.91

S31 --- OPEN. Separate options for running just TAP or just STAHET are to be created.
FIVR SIR25, WL1.2.92

C14 --- OPEN. Monitoring jobs using user 'ADP'.
FIVR SIR20, WL1.2.75

----- New Errors V1.2b -----

- S32 --- OPEN. File names used in side C-Mold to be mentioned in manual.
FIVR SIR35
- S33 --- OPEN. Incorporate resolution of FIVR SIR38 in the manual.
FIVR SIR39
- S34 --- OPEN. Improve FTPMAT data entry.
FIVR SIR40
- S35 --- OPEN. Invalid TAPPAT error message.
FIVR SIR41
- S36 --- OPEN. STAHET input deck title ends on second line.
FIVR SIR52
- S37 --- OPEN. Wrong message after ending C-Mold.
FIVR SIR55
- C15 --- OPEN. Parent window processes locked out.
FIVR SIR33
- C16 --- OPEN. C-Mold file name inconsistency.
FIVR SIR34
- C17 --- OPEN. STAHET_MAGNA needs to delete FIELD 1.
FIVR SIR44
- C18 --- OPEN. Existence of FIELD 1 causes STAHET_MAGNA to bomb.
FIVR SIR46
- C19 --- OPEN. STAHET_MAGNA applies pressures to nodes instead of faces.
FIVR SIR48
- C20 --- OPEN. Extraneous data in PATSTAHET menus.
FIVR SIR53
- C21 --- OPEN. ISMO not always 1.
FIVR SIR54
- C22 --- OPEN. Invalid C-Mold option.
FIVR SIR56
- E27 --- OPEN. STAHET control deck missing ALT. and MACH data.
FIVR SIR36
- E28 --- OPEN. FTPMAT is extremely unstable.
FIVR SIR37
- E29 --- OPEN. FTPMAT. The units conflict between analyses.
FIVR SIR38
- E30 --- OPEN. TAPPAT unable to process nodal data.
FIVR SIR42
- E31 --- OPEN. TAP_MAGNA only interpolate onto the "rim", and then is
unable to process the data.
FIVR SIR43
- E32 --- OPEN. STAHET_MAGNA error mapping because of streamlines.
FIVR SIR45
- E33 --- OPEN. STAHET_MAGNA doesn't work if other temps already exist.
FIVR SIR47
- E34 --- OPEN. No material data is put into the MAGNA input deck.
FIVR SIR49
- E35 --- OPEN. TAP_RENUMBER rennumbers elements wrong.
FIVR SIR50
- E36 --- OPEN. NSL not put into STAHET input deck.
FIVR SIR51

Summary of OPEN/Closed SIRs.

	S#	C#	E#	Total
Open	13	6	1	20
Pre V1.2b Closed	18	8	25	51
Totals				
Grand	31	14	26	71
New Problems V1.2b	6	8	10	24
Total Open Problems	19	14	11	44
Total Closed Probs.	17	8	25	51
Grand Total	37	22	36	95

Figures:

Figure 1. STAHET temperature results of the forebody in the region of the canopy.

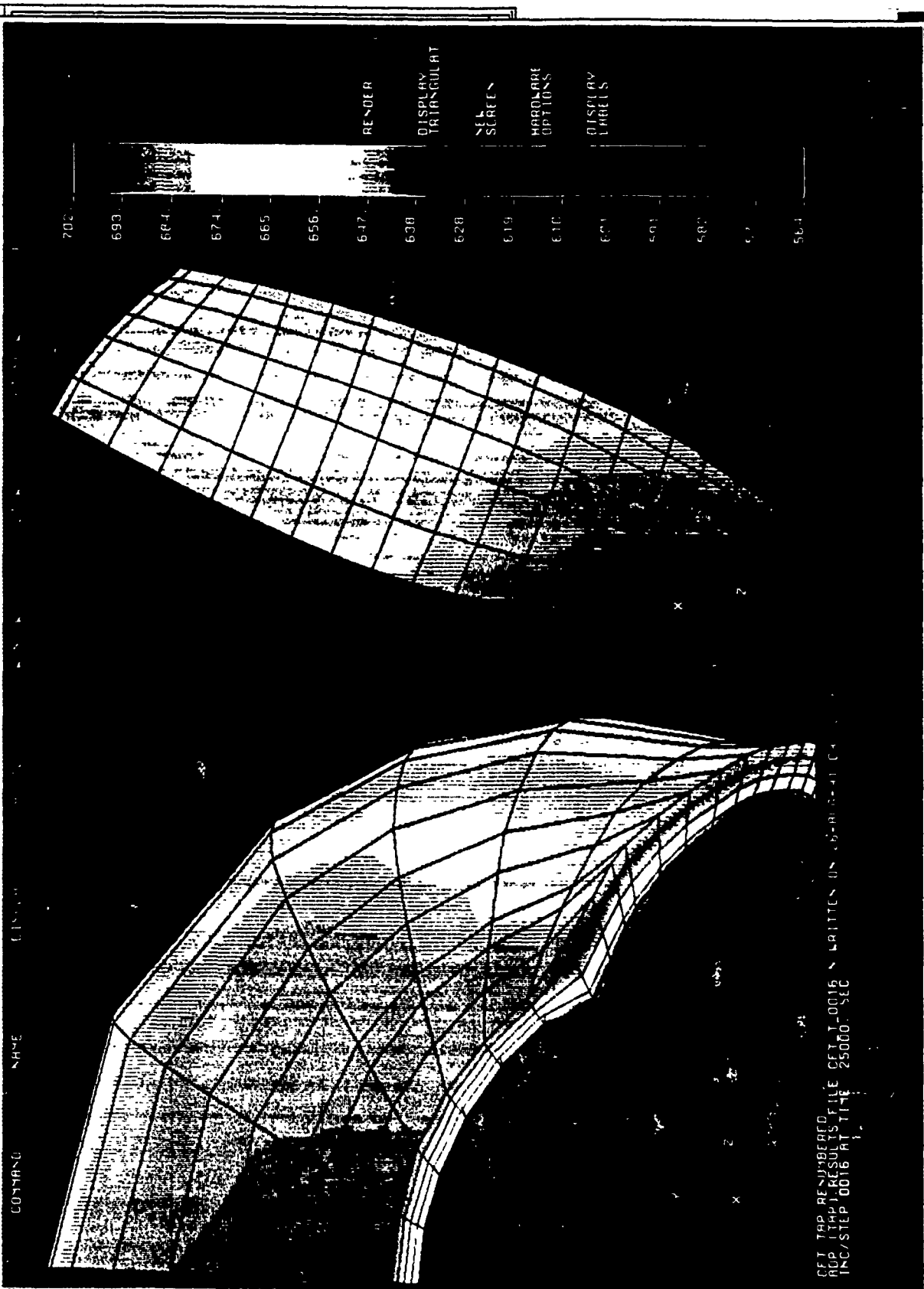
Figure 2. TAP temperature results of a canopy after a simple flight mission.

Figure 3. MAGNA static deflections due to applied STAHET pressures and TAP temperatures.

Figure 4. MAGNA frequency analysis results, the 2nd mode shape.

Figure 5. MAGNA birdstrike results. 0.001 seconds into the strike of a 4 pound bird at 500 knots.

PAIRAN24



CFI TOP RENDERED
 ROP LTPP RESULTS FILE CFI J 0016 N WRITTEN ON 48-HR-41 C4
 INC/STEP 0016 AT TIME 25000. SEC

Figure 2.

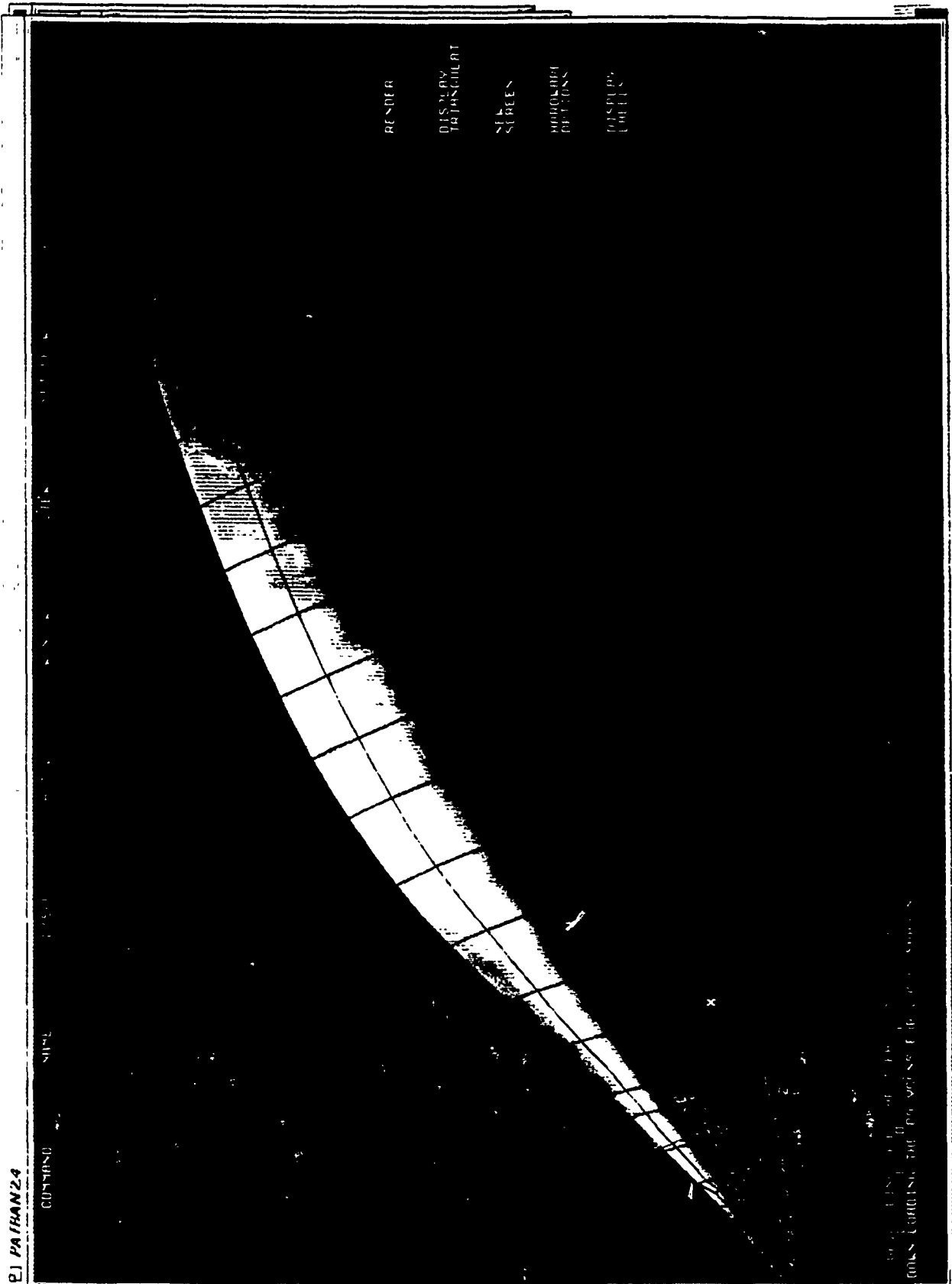


Figure 3.

PAIRAN24

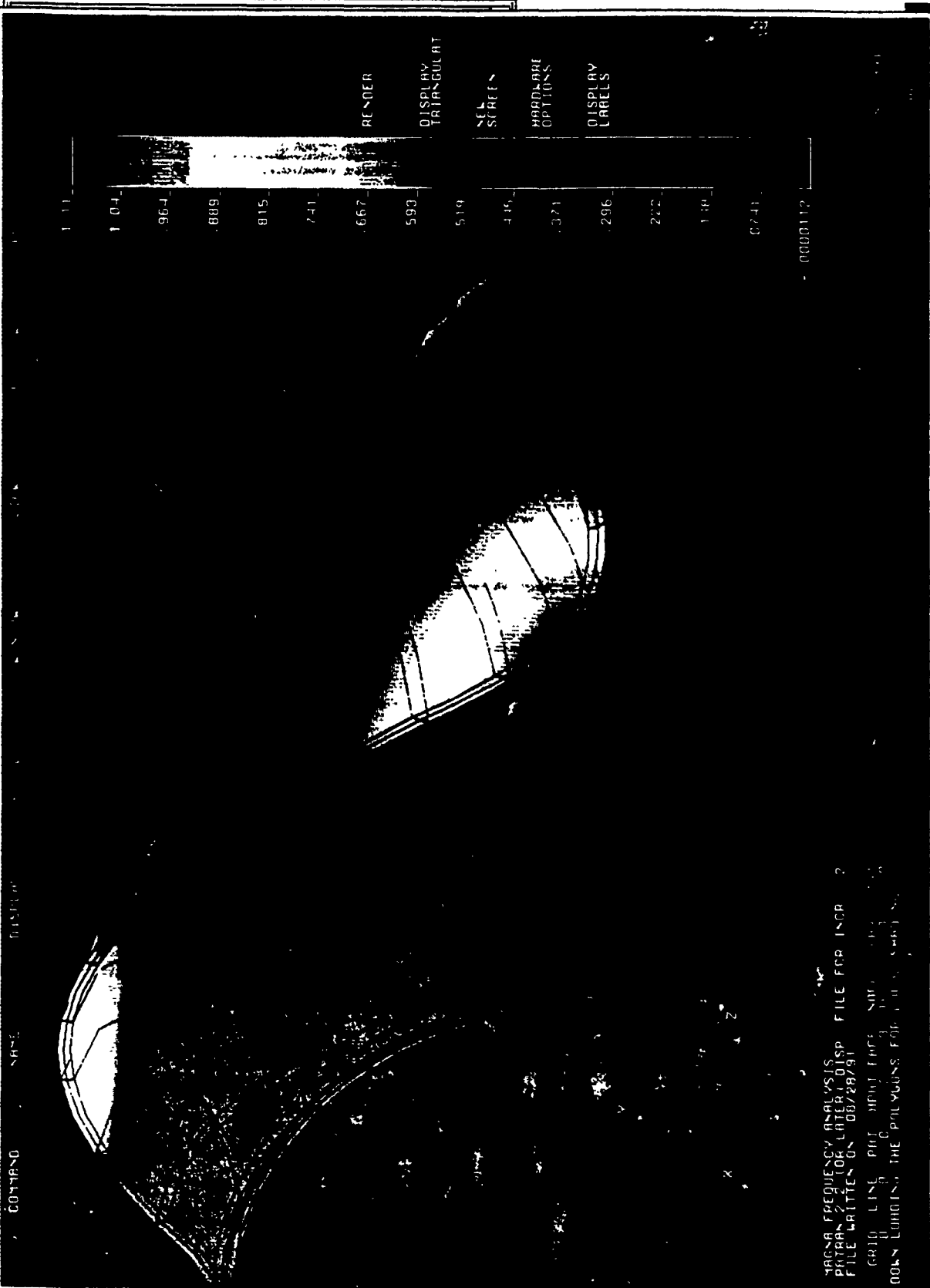
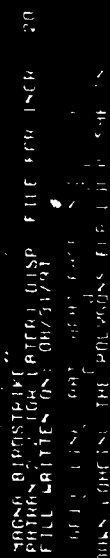


Figure 4.



25-19

EXPERIMENTAL INVESTIGATION OF THE INFLUENCE OF CONSTRAINED-LAYER DAMPING TREATMENT ON PARAMETRIC AND AUTOPARAMETRIC RESONANCES IN NONLINEAR STRUCTURAL SYSTEMS

Lawrence D. Zavodney
Assistant Professor

Joseph A. Schudt
Graduate Research Assistant

Department of Engineering Mechanics
The Ohio State University
209 Boyd Laboratory
155 West Woodruff Ave
Columbus, Ohio 43210

1.0 ABSTRACT

The influence of viscoelastic constrained-layer damping treatment on parametric resonances of single-degree-of-freedom (SDOF) systems and autoparametric resonances of multiple-degree-of-freedom (MDOF) nonlinear systems possessing internal resonance was investigated. The results show that commercially available aluminium-backed treatment is effective in suppressing parametric resonances of a particular mode; it essentially adds linear viscous damping to the system and moves the regions of parametric resonance away from the frequency axis. In the MDOF systems the damping affects the nonlinear coupling between modes and can suppress the modulation between modes. In general, its effect is to reduce the regions of nonlinear modal interaction and, in some cases, actually suppress it entirely with a sufficient amount of damping treatment. Experimental results include slow swept-sine excitations at constant amplitude and slow swept-amplitude excitations at constant frequency. Particular attention was paid to the nonlinear resonances and the modal interaction regions bounded by the Hopf bifurcation.

2.0 INTRODUCTION

Parametric resonances are not uncommon in structural vibrations. Zavodney¹ and Zavodney, et al²⁻⁴ have provided a summary of parametric resonance in SDOF nonlinear structural systems. Damping is a common method of controlling or reducing vibration for externally excited resonances. However, damping plays quite a different role in parametrically excited systems. When systems exhibit nonlinear behavior, the analysis is more complicated; the nonlinearity affects the system response along with the damping, so it is not always obvious what part the damping plays in the response. Zavodney and Shihada⁵ investigated the influence of the linear viscous damping coefficient on the fundamental and principal parametric resonances of SDOF systems possessing quadratic and cubic nonlinearities. The results showed that linear viscous damping plays a significantly different role; if reduction in amplitude at resonance is the objective, then in some cases certain critical levels of damping must be exceeded--otherwise one can increase the damping coefficient by an order of magnitude but realize less than 5% reduction in the response amplitude.

When the structure possesses more than one mode, as is usually the case, there is the possibility of nonlinear modal coupling. When modes are coupled, it is possible to exchange energy from a directly excited mode to another mode. The end result is that more than one mode is participating in the response and, hence, the structure is vibrating at other frequencies in addition to the frequency of the excited mode--which is not always the same as the excitation frequency. As a result the analysis becomes even more complicated. Attempting to find appropriate mathematical models for these behaviors is extremely difficult. One particular mathematical model may describe one type of behavior, but when the excitation frequency is changed by only 0.1 Hz, the behavior is something else--qualitatively and quantitatively--i.e., a bifurcation has occurred. The

nonlinear coupling responsible for this behavior is further enhanced if the structure possesses an internal resonance; an internal resonance occurs whenever any natural frequencies are commensurate (i.e., in a 2:1, 3:1, 3:2, etc. ratio). When a structure simultaneously possesses an internal resonance and appropriate nonlinear coupling terms, it is possible for one mode to parametrically excite another mode; this phenomenon is called an autoparametric resonance. When this happens, the nonlinear effects are greatly intensified and completely dominate the response. Nayfeh and Zavodney⁶ and Balachandran and Nayfeh⁷ provided a theoretical model and experimental results showing that such behavior can lead to long-time responses that are not steady-state; mathematically a Hopf bifurcation has occurred.

The objective of this project was to conduct experiments on nonlinear SDOF and MDOF flexible structures using the experimental facilities at Wright Laboratory to study the effects of damping on parametric and autoparametric resonance. The experiments were performed on structures fabricated from uniform prismatic beams and lumped masses. These types of structures were easy to prepare and tune--i.e., by adjusting the position of the masses the resonant frequencies could be changed. This was essential for the MDOF structure when an internal resonance was desired.

3.0 SINGLE-DEGREE-OF-FREEDOM STRUCTURE

Experiments were performed on a flexible beam carrying a lumped mass; this structure is shown in Figure 1. The excitation was a base displacement (along the axis in the vertical direction) at a frequency twice that of the first flexural mode of the cantilever beam. The resonant response that ensues is called a principal parametric resonance. The experiments consisted of frequency sweeps at constant-amplitude acceleration and amplitude sweeps at constant frequency. The amplitude was held constant by a computer-controlled feedback loop; as the

frequency was changed, the corresponding excitation amplitude required adjustment to keep the acceleration level constant. The excitation level was chosen as large as possible without causing excessive amplitudes of displacement; it turned out that a level of 0.350 g's was selected. The response was measured by strain gages mounted at the base of the lower beam.

The frequency response of the structure before application of the damping treatment for two levels of excitation is shown in Figure 2. It shows that the system is softening because it bends to the left. Zavodney¹ and Zavodney and Nayfeh⁸ performed experiments on similar structures and found responses that exhibited hardening behavior. It turns out that the relative length (i.e., flexibility) of the beam determines, in part, whether the system is hardening or softening. These experiments were conducted by increasing the frequency of excitation very slowly while simultaneously keeping the table acceleration constant. The arrows indicate jumps; for example, when the excitation level is 0.350 g's (denoted by the circles), the response jumps up to the large amplitude and then slowly decreases as the frequency is increased. When the frequency is decreased, the response follows the same curve but extends it somewhat. The response eventually jumps down to the trivial response. The response for the 0.250-g excitation level is qualitatively the same but at a lower amplitude. Additional experiments were performed with one strip and two strips of viscoelastic damping treatment as shown in Figures 3 and 4, respectively. For the lower excitation level, the parametric resonance was completely suppressed using only two strips of damping treatment. Two 0.10-inch strips of damping treatment cover only 5% of the surface area of the beam between the support and the lumped mass. In summary, increased application of the damping treatment reduces the amplitude of the resonant response, and reduces the region of parametric instability.

The regions of parametric resonance in the excitation-amplitude versus the excitation-frequency plane were also determined. These experiments were

performed by repeating the frequency response experiments for different levels of excitation and plotting only the bifurcation points where either the trivial response becomes unstable (during both a sweep up and a sweep down) or a nontrivial response jumps to the trivial response. An example is shown in Figure 5. This figure shows three boundaries; the inner-most region (bounded by the two curves on the right) represents the loss of trivial stability for the linear system; the additional curve on the far left represents the extension of the instability region caused by the overhang due to the softening nonlinearity in the system. Because nontrivial responses exist at excitation frequencies below those predicted by linear theory, the system is said to possess a subcritical instability. This overhang is reduced for increased damping levels as shown in Figures 6 and 7.

A summary of the bifurcation boundaries showing only the region where nontrivial responses exist is shown in Figure 8. As the equivalent linear viscous damping coefficient increases, the instability regions move away from the frequency axis. From this figure one can see that the only way to completely suppress a parametric resonance is to shift the instability region far enough away from the axis so that it is completely removed from the area of interest.

4.0 MULTIPLE-DEGREE-OF-FREEDOM STRUCTURE

Experiments were performed on a multiple-degree-of-freedom (MDOF) structure. The objective was to study the effect of viscoelastic damping treatment on the autoparametric resonance (modal interaction) caused by the simultaneous occurrence of appropriate nonlinear coupling terms with an internal resonance. The structure used for these experiments is shown in Figure 9; it has been studied by several researchers^{7,9,10}. It was chosen because it was easy to fabricate and could be easily tuned (to create an internal resonance) by adjusting the lengths of the beams and position of m_2 . The procedure for applying damping treatment to the SDOF structure was used for the MDOF structure: thin strips of 0.10-inch

width damping treatment were applied in increasing numbers to each beam, effectively increasing the damping of the structure.

The responses of this structure--both to impulse and harmonic excitation--are very complicated. The primary source of the quadratic modal-coupling nonlinearity in this structure is the asymmetrical geometry. When an internal resonance exists, this nonlinear coupling is significantly intensified and can dominate the response; when this happens the response amplitudes can be three or four times larger than the linear response. This happens whenever one mode is (directly or parametrically) excited; the response initially consists of the excited mode--which is the linear response. Because of the internal resonance (in this case it appears as a 2:1 frequency ratio between the first and second mode), whenever the second mode is excited, it acts as a parametric excitation to the first mode. As a result, when appropriate conditions exist, the first mode will also be excited. Because it is a parametric excitation, the response requires many cycles to achieve a steady state.

The purpose of applying the damping treatment was to eliminate (or at least attenuate) the large amplitude responses that are caused by nonlinear modal coupling (e.g., saturation and modulation). Hence, to ascertain these effects, both frequency sweeps at constant amplitude and amplitude sweeps at constant frequency were performed for three levels of damping treatment. The labels correspond to the following amounts of damping treatment: the unmarked (or back plane) for none, "I" for two strips on each beam (17% of the surface area), "II" for six strips on each beam (50%), and "III" for 100% coverage.

The experiments were performed at very high levels of excitation. Previous studies^{6,7,9,10} have shown that under some conditions, nonlinear responses can be achieved with very small levels of excitation (on the order of 20 mili-g's). In the experiments performed here, the excitation levels were on the order of 100 mili-g's--five times that required to solicit a nonlinear response consisting of coupled

modes. This large excitation level was used to extend the range for which damping treatment could be applied; in other words, more damping treatment could be applied before the nonlinear motion was expected to be suppressed.

The structure shown in Figure 9 was tuned such that the first resonance occurred at 4.012 Hz and the second resonance occurred at 8.040 Hz; this corresponds to an internal resonance detuning of +0.016 Hz, or +0.20% (of the second resonant frequency). This was as close as feasibly possible to perfect tuning on a real structure; a tuning closer to 0% could be accomplished at the expense of more painstaking effort (i.e., moving the lower beam into the clamped support in 0.001-inch increments and repeating the experiments until the desired detuning is achieved). After each application of damping treatment, the model was retuned to get as close as possible to 0.0%. Experiments for a slightly detuned structure were also performed; the detuning was accomplished by increasing the mass of m_2 (by adding a small screw). In this case, the resonances occurred at 3.750 Hz and 7.750 Hz, which corresponds to a detuning of +0.25 Hz, or +3.2%. The results of these experiments are qualitatively the same, but the increased internal-resonance detuning causes skewing of the frequency-response curves and increases the region where modulated responses occur. Due to the length restrictions of this report, these results are not discussed here.

The undamped structure had the largest response levels. The frequency responses for the structure without and with three amounts of damping treatment are presented in one figure; this 3-D perspective viewing aids in seeing the qualitative and quantitative changes caused by the increased damping. When the first mode is directly excited with a high-level excitation, the response is nonlinear, as shown in Figure 10. Figure 10(a) shows the first mode response and Figure 10(b) shows the second mode response. The frequency axis is shown normalized with respect to the excitation frequency (because the addition of the

damping treatment caused the frequencies to shift slightly) and the response amplitude (displacement of m_1) is represented in units corresponding to the ratio of the lower mass displacement amplitude to the length of the lower beam. This scaling provides a convenient comparison of the results of all of the experiments. For example, the peak amplitude of the response of the first mode is approximately 8% of the length of the lower beam.

The response of the untreated structure shows the most interesting behavior. As the frequency of excitation is increased from 0.90, the amplitude initially grows until the second mode is excited; this happens at a frequency near 0.98. As the frequency is increased further, the amplitude of the first mode decreases dramatically--almost four times! When a Hopf bifurcation occurs near a frequency of 1.0 no steady-state responses are possible. This region is denoted by dots which represents an "average" amplitude. Actual variations are shown later in the amplitude response curves shown in Figure 12. As the frequency is further increased, the amplitude of response increases until it jumps down (at the first mode frequency of 1.072) to the linear response; the linear response consists only of the directly excited mode--no other modes are present. If the frequency is decreased from above, the jump up occurs at a lower frequency (i.e., 1.044) than did the jump down during a sweep up. This indicates an overhang or double-valued steady-state response. Theory shows that these two solution branches are connected by an unstable solution branch; this is shown by the dashed line on the response (note--there are no data points for this dashed line because they cannot be realized in the laboratory). Further decreases in the excitation frequency cause the response to follow the same path as that followed during the sweep up. The frequency response shown in this curve is almost symmetrical; because it is slightly skewed to the right, it indicates slight positive detuning of the internal resonance.

In summary, the frequency response curve shows four distinct types of motions that are possible when the excitation frequency is near the first mode: (1) the linear solution where only the directly excited mode participates in the response, (2) a region where nonlinear coupling is present and causes two steady-state solutions to exist--the linear one and a nonlinear one, (3) a nonlinear response where the modes achieve steady-state amplitudes, and (4) a region where no steady-state solutions exist.

As damping treatment is applied to the structure, response amplitudes are attenuated. However, certain trends can be observed in Figure 10. The first application of damping treatment consisted of two strips on each beam (17%). Both response peaks are broadened (indicating increased damping); the peak at the lower frequency is almost eliminated while the response peak at the upper frequency is reduced in amplitude sufficiently to eliminate the hysteresis. The second application of damping treatment (six strips, 50%) causes further peak smoothing. The third application (100%) almost completely eliminates the modal coupling. In summary, the damping treatment can attenuate the modal responses, and it appears that a sufficient quantity could eliminate the nonlinear coupling altogether.

A direct excitation to the second mode is shown in Figure 11. Similar observations can be made from these curves shown in a 3-D perspective plot. The three curves correspond to the same amounts of damping treatment used for the direct excitation to the first mode shown in Figure 10. In Figure 11 the frequency axes correspond to a nondimensionalization with respect to the excitation frequency; hence the second mode has a resonance at 1.0 and the first mode, representing a frequency of one-half that of the excitation, has a resonance at 2.0. The indirectly excited mode (i.e., the first mode) remains trivial until it is strongly (and nonlinearly) coupled to the second mode; when it is, the amplitude of the second mode is drastically attenuated. Furthermore, during a sweep down,

there is no jump up in the second mode response--only a jump down--because the lower branch merges with the upper branch; it is not a turning point bifurcation as was the case for a direct excitation to the first mode as seen in Figure 10. The first mode demonstrates both a jump up and a jump down during a frequency sweep. For the third application of damping treatment, the nonlinear coupling between the modes is completely suppressed.

Experiments were also conducted to measure the amplitude response at selected frequencies of excitation; these curves are cuts across a particular frequency response curve shown in Figure 10. For the case of no damping treatment, a cut at a frequency of 1.004 passes through the modulation region. The amplitude response at this frequency is shown in Figure 12. As the amplitude of excitation is increased, the system experiences a Hopf bifurcation and begins to modulate; i.e., energy begins to flow back and forth between the two modes. As the amplitude of excitation is increased, the mean tends to increase and the maximum excursions increase.

At a frequency of 1.067 the frequency response curve (for no damping treatment) in Figure 10(a) shows a double value. An amplitude sweep through this region, as shown in Figure 13, reveals a jump phenomenon. When starting on the lower branch, increasing the amplitude of excitation causes a jump up to the upper branch; decreasing the amplitude of excitation causes a jump down at a lower value of the amplitude of excitation.

When the maximum damping treatment was applied, there was essentially one qualitative response; Figure 14 shows an amplitude sweep through the resonance at a frequency of 0.979. Although the curves are quite tame, they show nonlinear coupling because the second mode is excited. Even though the excitation is driving the first mode, some of the energy is channeled into the second mode.

Amplitude excitation experiments were also performed for a direct excitation to the second mode. Figure 15 shows a cut across the 0.998 frequency line (the 1.999 frequency line for the first mode) in Figure 11(b) for the structure with no damping treatment applied. Although the second mode is directly excited, its response is attenuated; instead the energy goes into the first mode which responds like a parametrically excited SDOF system. An amplitude sweep at the 1.057 frequency line (2.119 for the first mode) in the region of the overhang is shown in Figure 16. During the experiment, the amplitude of the excitation was increased slowly. At 0.32 g's rms, the linear solution becomes unstable. However, the divergence is so slow that it is possible to stay on the unstable solution long enough to locate the equilibrium points. These are shown in the figure as circles without a solid line connecting them. However, by waiting long enough, the response goes to the steady-state nonlinear solution which is shown with the solid line. During a sweep down, the response follows the upper curve rather than the lower curve. At an excitation level of 0.075 g's rms, the response jumps down to the linear response on the lower branch and remains there. Note that the overhang region in Figure 11 is qualitatively different than the overhang region in Figure 10; hence, it is not surprising to see a qualitative difference in the double solution region of Figures 13 and 16.

When damping treatment was applied to the structure, the regions of multiple solutions contracted and eventually disappeared. For small levels of excitation the linear solution appears, as shown in Figure 17. At this frequency of excitation (0.992 for the first mode, 1.991 for the second mode), when the excitation level exceeds 0.09 g's, the second mode saturates. The energy that is put into the system at the second-mode natural frequency is passed into the first mode; essentially the second mode is parametrically exciting the first mode. Hence, the first mode response appears to have a parametric type response.

5.0 SUMMARY

A series of experiments were performed to study the effect of commercially available viscoelastic damping treatment on parametric and autoparametric resonances in nonlinear systems. Both SDOF and MDOF structures were used; they consisted of flexible beams and masses. For the SDOF structure, experiments were conducted at several levels of excitation and it was found that the damping treatment was particularly effective in reducing and even suppressing the resonance entirely. For the 2DOF system, the particular case of an internally resonant structure was considered. The qualitative and quantitative effects of the damping treatment were determined for very large excitation levels; it was observed that the damping treatment first reduces the nonlinear modal coupling and modulation regions and then reduces the amplitude of response. For the case of direct excitation to the second mode, damping treatment was capable of eliminating the nonlinear coupling, even when the internal resonance detuning was on the order of 0.20%.

6.0 REFERENCES

1. Zavodney, L.D., "A Theoretical and Experimental Investigation of Parametrically Excited Nonlinear Mechanical Systems," Ph.D. Dissertation, Virginia Polytechnic Institute and State University, 1987.
2. Zavodney, L.D. and Nayfeh, A.H. "The Response of a Single-Degree-of-Freedom System with Quadratic and Cubic Nonlinearities to a Fundamental Parametric Resonance," *Journal of Sound and Vibration*, 1988, Vol 120, 63-93.
3. Zavodney, L.D., Nayfeh, A.H. and Sanchez, N.E., "The Response of a Single-Degree-of-Freedom System with Quadratic and Cubic Nonlinearities to a Principal Parametric Resonance," *Journal of Sound and Vibration*, 1989, Vol 129, 417-442.
4. Zavodney, L.D., Nayfeh, A.H. and Sanchez, N.E., "Bifurcations and Chaos in Parametrically Excited Single-Degree-of-Freedom Systems," *Nonlinear Dynamics*, 1990, Vol 1, 1-21.
5. Zavodney, L.D. and Shihada, S.M., "The Role of Damping in the Suppression of Parametric Resonances in Nonlinear Systems," *Proceedings of Damping '89*, West Palm Beach, Florida, February 8-10, 1989, FBD-1 - FBD-22.
6. Nayfeh, A.H. and Zavodney, L.D., "Experimental Observation of Amplitude- and Phase-Modulated Responses of Two Internally Coupled Oscillators to a Harmonic Excitation," *Journal of Applied Mechanics*, 1988, Vol 110, 706-710.
7. Balachandran, B. and Nayfeh, A.H., "Nonlinear Motions of Beam-Mass Structure," *Nonlinear Dynamics*, 1990, Vol 1, 39-62.
8. Zavodney, L.D. and Nayfeh, A.H., "The Non-Linear Response of a Slender Beam Carrying a Lumped Mass to a Principal Parametric Excitation: Theory and Experiment," *International Journal of Nonlinear Mechanics*, 1989, Vol 24, 105-125.
9. Zavodney, L.D. and Hollkamp, J.J., "Experimental Identification of Internally Resonant Nonlinear Structures Possessing Quadratic Nonlinearity," *Proceedings of the 32nd Structures, Structural Dynamics and Materials Conference*, Baltimore, Maryland, 8-10 April, 1991, 2755-2765.
10. Haddow, A.G., Barr, A.D.S., and Mook, D.T., "Theoretical and Experimental Study of Modal Interaction in a Two-Degree-of-Freedom Structure," *Journal of Sound and Vibration*, 1984, Vol 97, 451-473.

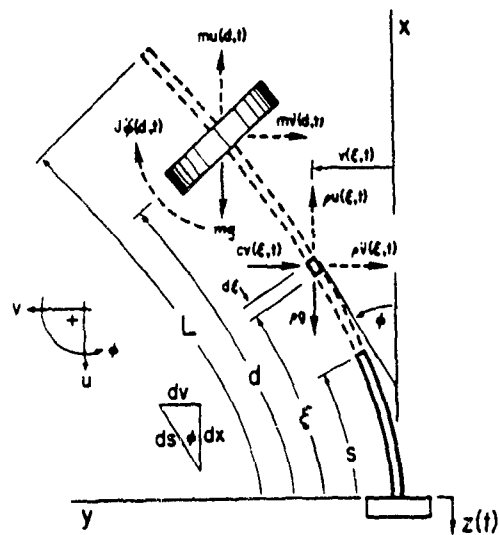


Figure 1. Structure consisting of a flexible beam carrying a lumped mass used for the SDOF experiments. Constrained-layer damping treatment was applied in thin 0.10-inch strips on both sides to incrementally increase the level of damping in the structure. The model was attached to a shaker head that was oscillating in the vertical direction.

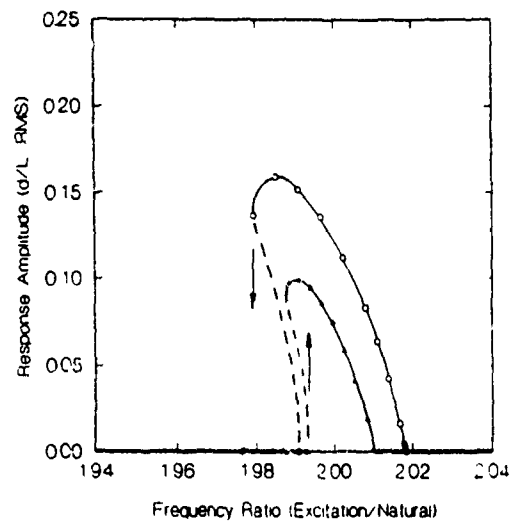


Figure 3. Frequency-Response at two levels of excitation for the SDOF structure with one strip of damping treatment applied.

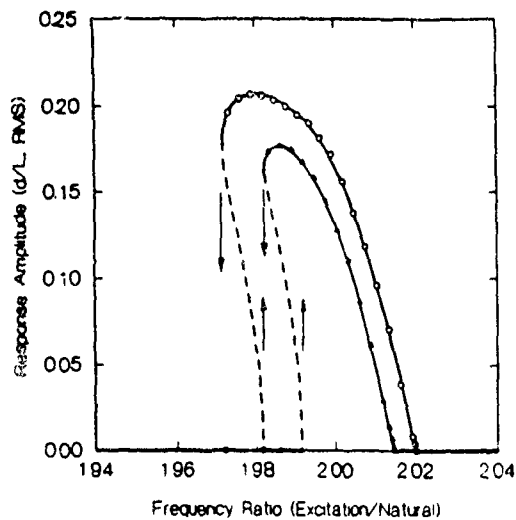


Figure 2. Frequency-Response for the SDOF structure before damping treatment was applied for two levels of excitation (0.25 and 0.35 g's rms). The bending of the curves to the left indicates that the system is softening. These curves were obtained by sweeping the frequency up and down through the region of principal parametric resonance at a rate slow enough such that a steady-state amplitude was maintained.

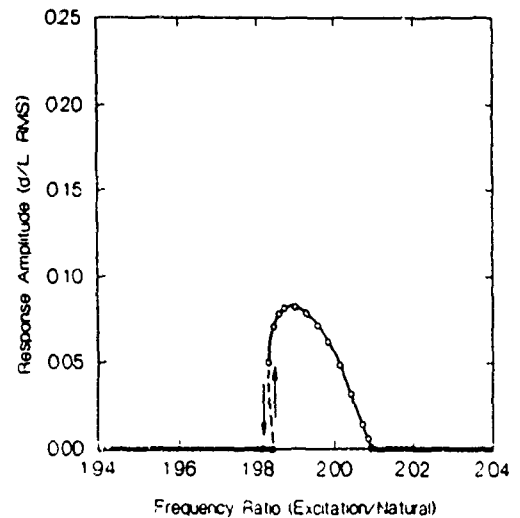


Figure 4. Frequency-Response at two levels of excitation for the SDOF structure with two strips of damping treatment applied. Note that the resonance for the smaller excitation level is completely suppressed.

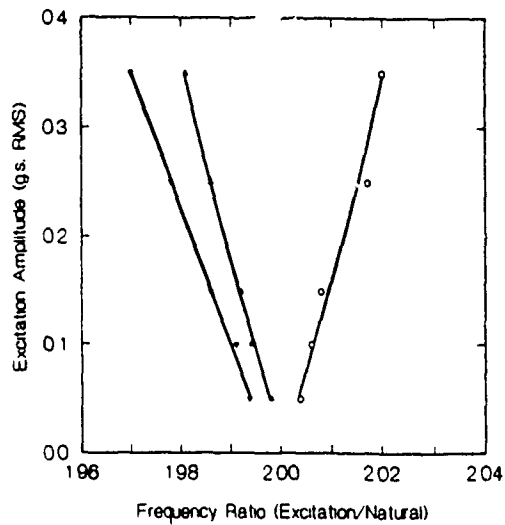


Figure 5. Bifurcation diagram showing the instability regions of the principal parametric resonance for the beam before the damping treatment was applied. The curve on the far left represents the bifurcation boundary caused by the overhang observed in Figure 2 which is caused by the softening behavior. It represents a subcritical instability.

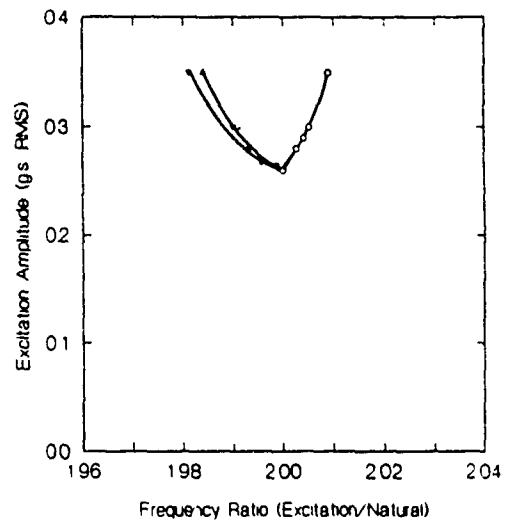


Figure 7. Bifurcation diagram showing the instability regions of the principal parametric resonance for the beam with two strips of damping treatment.

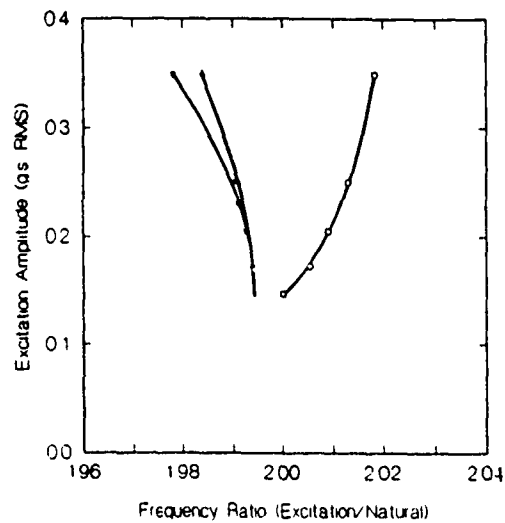


Figure 6. Bifurcation diagram showing the instability regions of the principal parametric resonance for the beam with one strip of damping treatment.

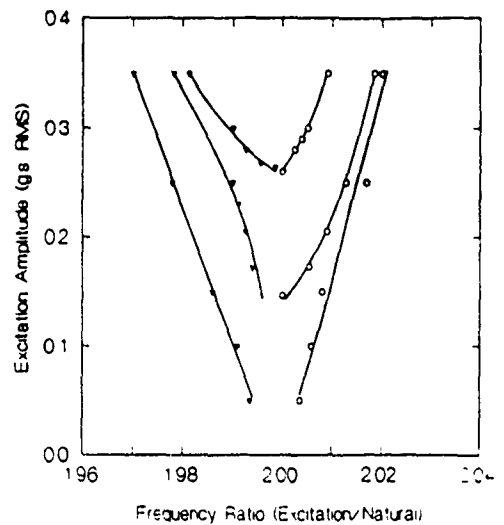


Figure 8. Bifurcation diagram summarizing the region where nontrivial responses exist for three levels of damping. Increased damping treatment causes the instability region to migrate away from the frequency axis and contract.

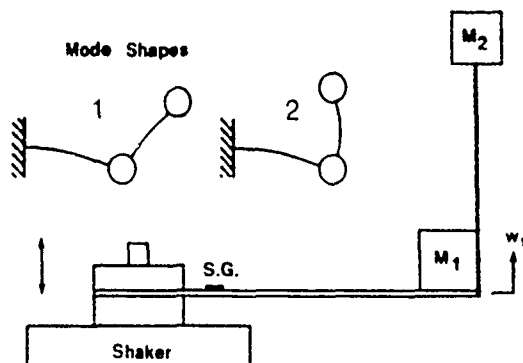


Figure 9. Structure used for the MDOF experiments. By adjusting the length of the lower beam and the position of the second mass, the first and second resonant frequencies could be adjusted. The higher modes and out-of-plane modes were not excited during the experiments; hence a 2DOF model can be used to adequately describe the observed behavior.

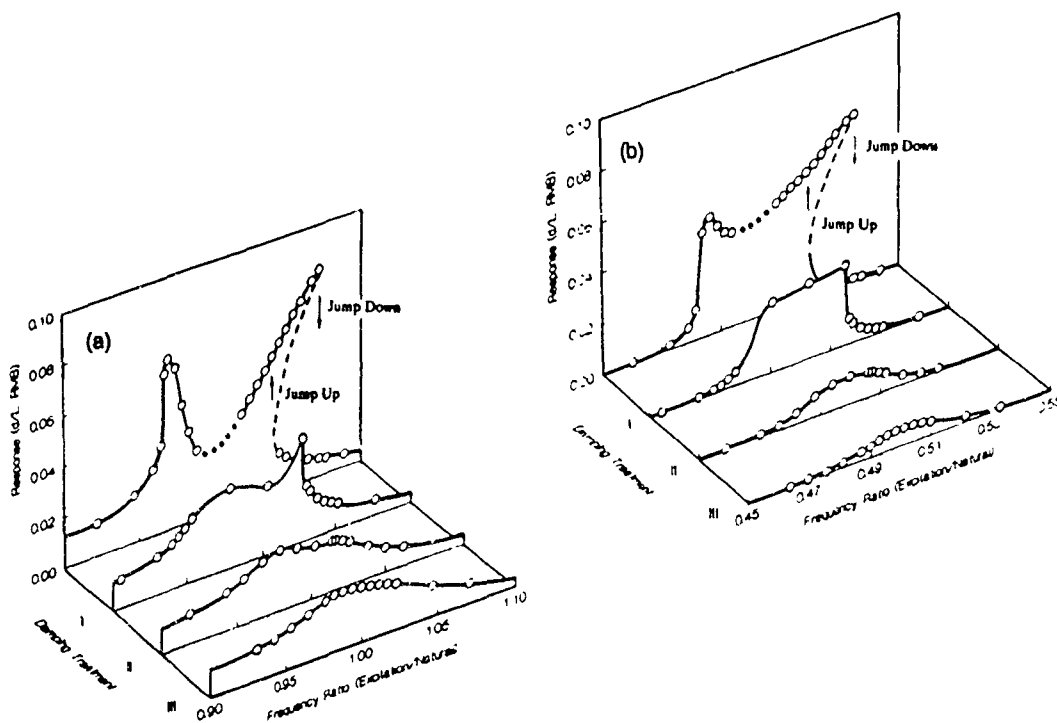


Figure 10. Frequency response of the 2DOF structure with +0.20% detuning (for no damping treatment) for a direct excitation to the first mode: (a) shows the first mode response and (b) shows the second mode response. Although attempts were made to have a perfect internal resonance, there is some small amount of detuning as evidenced from the unsymmetrical nature of the response curves.

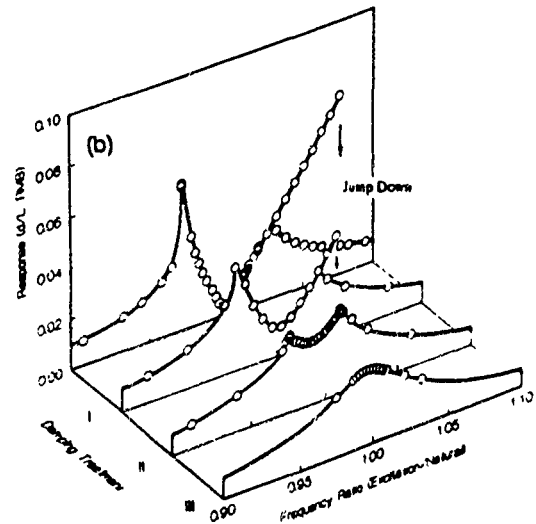
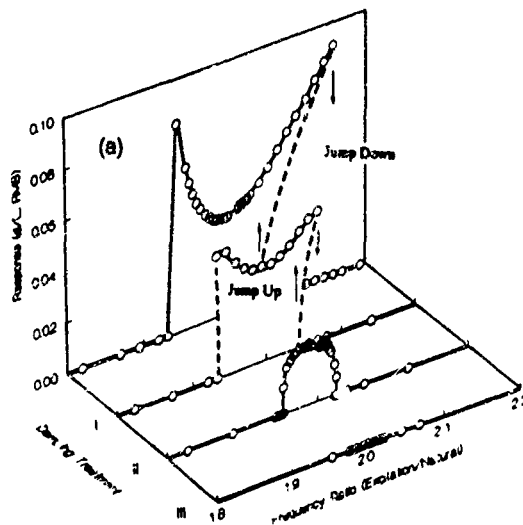


Figure 11. Frequency response of the 2DOF structure with +0.20% detuning (for no damping treatment) for a direct excitation to the second mode: (a) shows the first mode response and (b) shows the second mode response.

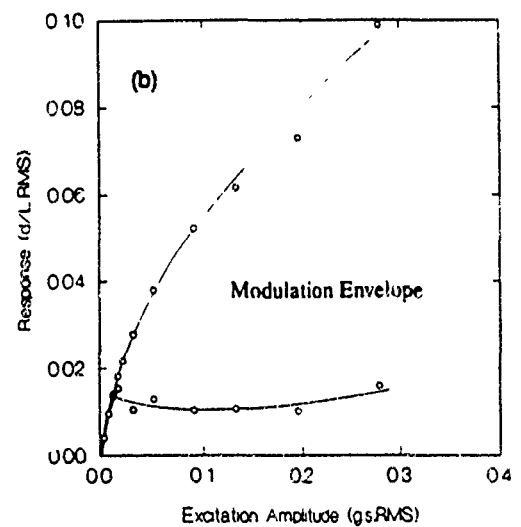
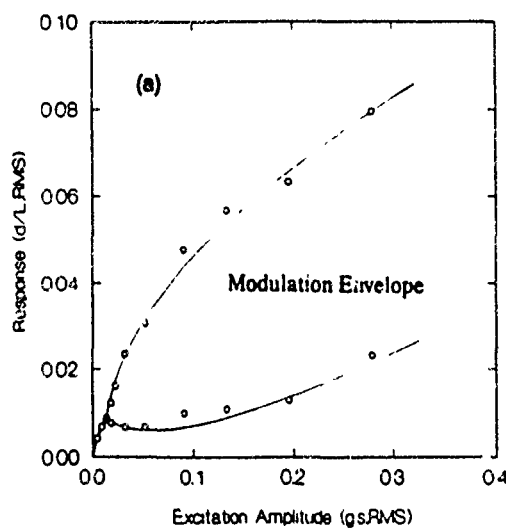


Figure 12. Amplitude response of the 2DOF structure with +0.20% detuning and no damping treatment for a direct excitation to the first mode at a frequency of 1.004. This frequency corresponds to the modulation region in Figure 10. The curves approximate the bounds on the modulation: (a) first mode amplitude, (b) second mode amplitude.

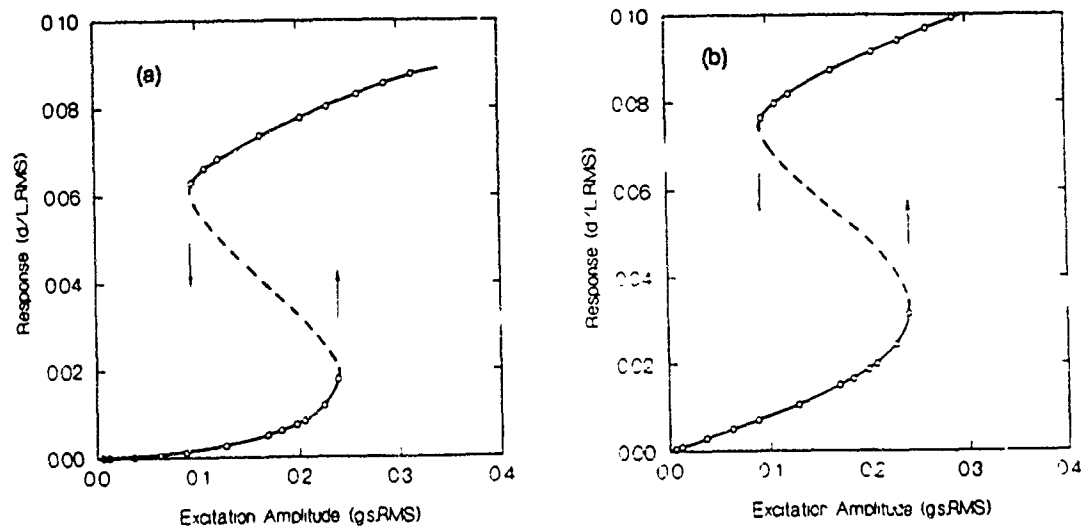


Figure 13. Amplitude response of the 2DOF structure with +0.20% detuning and no damping treatment for a direct excitation to the first mode at a frequency of 1.067. This frequency corresponds to the overhang region in Figure 10. The curves approximate the bounds on the modulation: (a) first mode, (b) second mode.

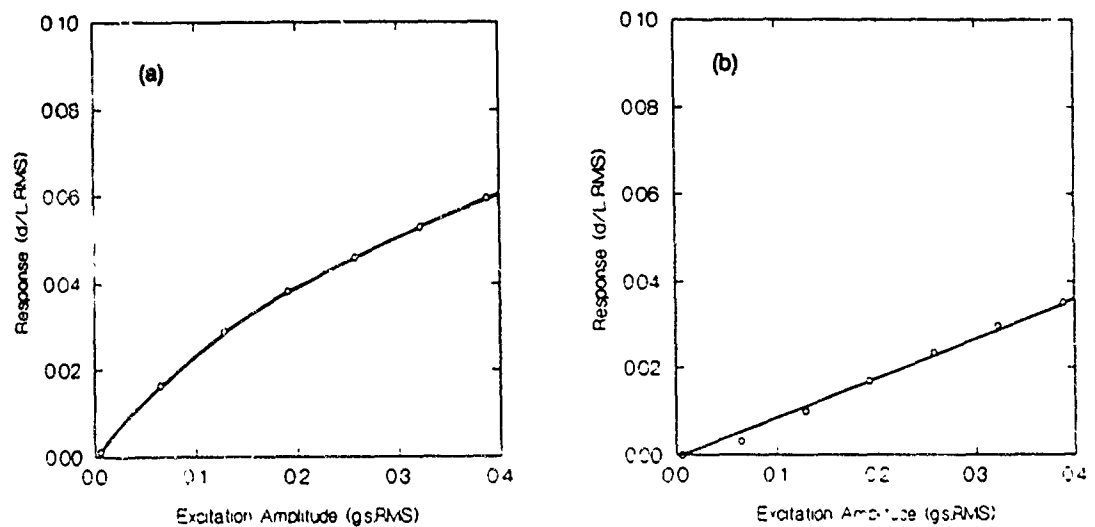


Figure 14. Amplitude response of the 2DOF structure with +0.29% detuning for the highly damped case for a direct excitation to the first mode at a frequency of 0.979. This frequency corresponds to the resonance region in Figure 10. The curves approximate the bounds on the modulation: (a) first mode, (b) second mode. These curves should be compared to Figure 12 to see the effect of the damping treatment—it completely suppresses the modulation and attenuates the amplitude of the response.

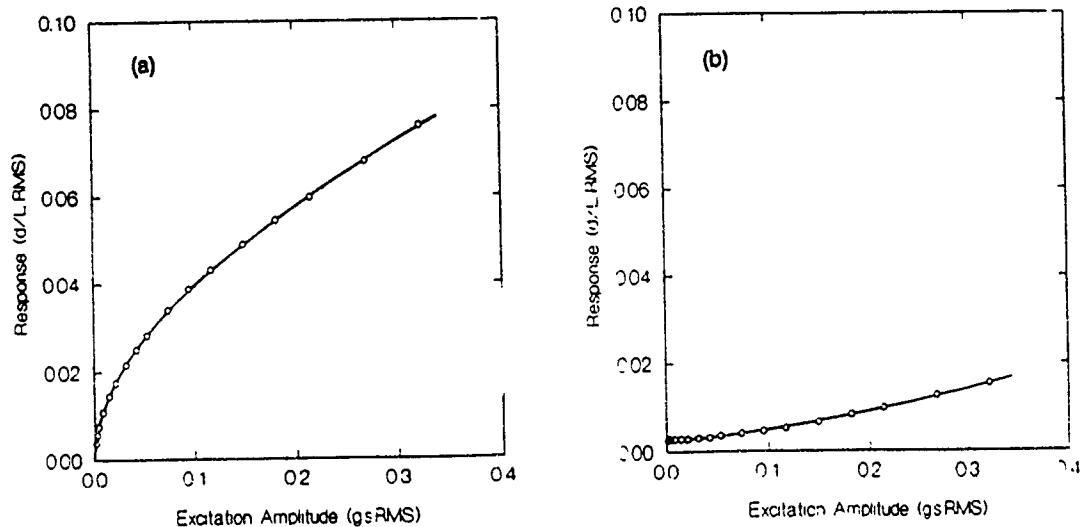


Figure 15. Amplitude response of the 2DOF structure with +0.20% internal resonance detuning for the case with no damping treatment for a direct excitation to the second mode at a frequency of 0.998. This frequency corresponds to the resonance region in Figure 11(b): (a) first mode, (b) second mode.

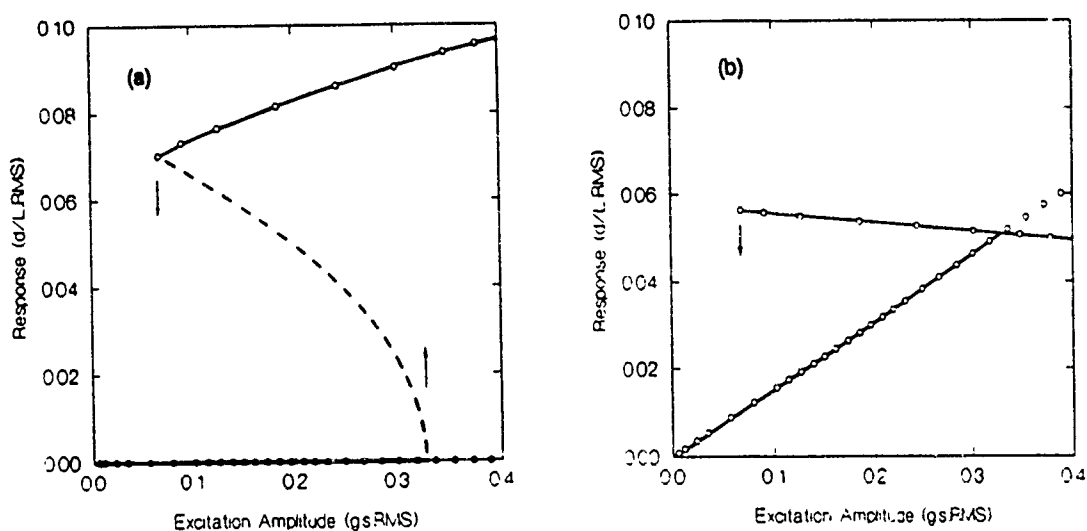


Figure 16. Amplitude response of the 2DOF structure with +0.20% internal resonance detuning for the case with no damping treatment for a direct excitation to the second mode at a frequency of 1.057. This frequency corresponds to the overhang region in Figure 11(b): (a) first mode, (b) second mode.

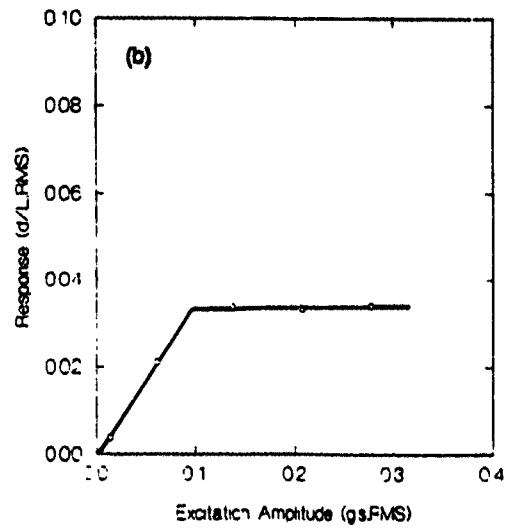
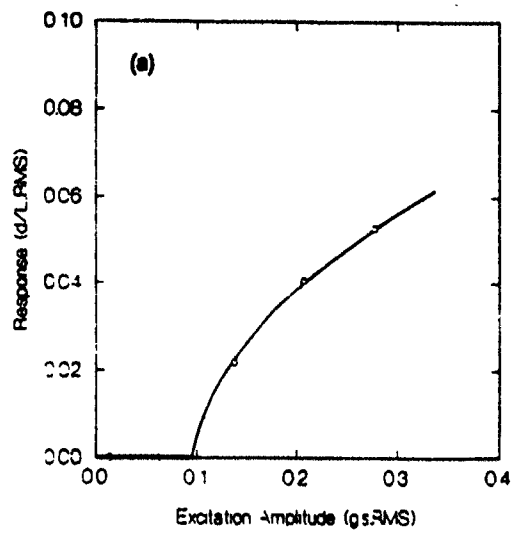


Figure 17. Amplitude response of the 2DOF structure with +0.29% internal resonance detuning for the highly damped case for a direct excitation to the second mode at a frequency of 0.992. This corresponds to the peak response in Figure 11: (a) first mode, (b) second mode.

1991 SUMMER RESEARCH PROGRAM

Sponsored by

UNITED STATES AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by

RESEARCH & DEVELOPMENT LABORATORIES

**APPLICATION OF COMPUTATIONAL FLUID DYNAMICS FOR OPTIMAL
STRUCTURAL DESIGN OF FLIGHT VEHICLES**

Prepared by:	Anne M. Stephenson
Academic Rank:	Graduate Assistant
Department & University:	Aerospace Engineering University of Dayton
Research Location:	Flight Dynamics Laboratory Structures and Dynamics Division Analysis/Optimization Branch
Focal Point:	Dr. Vipperla B. Venkayya
Date:	30 September 1991
Employee Number:	353

APPLICATION OF COMPUTATIONAL FLUID DYNAMICS FOR OPTIMAL
STRUCTURAL DESIGN OF FLIGHT VEHICLES

Anne M. Stephenson
Graduate Student

Aerospace Engineering Department
University of Dayton, Dayton, OH 45469-0227

ABSTRACT

This investigation concentrates on the use of both linear and non-linear aeroelastic models for the structural weight optimized design of a fighter-type vehicle. ASTROS produces the optimized weight redesign of a finite element model of the wing and determines a flutter boundary using both subsonic and supersonic aerodynamic theories. The flutter phenomenon is recreated by the non-linear transonic code XTRAN3S. Once a critical flutter point is identified, one of two approaches can be taken to correct the problem. Either a control system is devised to suppress flutter or the structure is redesigned to preclude flutter. In either case, a linear transonic aerodynamic model greatly facilitates the design process. The purpose of this study is to extract the linear transonic aerodynamic model.

ACKNOWLEDGEMENT

The following people contributed to a worthwhile, educational summer of research sponsored by the Air Force Office of Scientific Research and Research and Development Laboratories at the Flight Dynamics Directorate, Wright Patterson Air Force Base. Dr. Vipperla B. Venkayya offered counseling and expertise, fulfilling his roll as mentor. Dr. Franklin E. Eastep provided the opportunity for the summer research experience, as well as the topic of research.

Dr. Max Blair extended much time and theoretical guidance throughout the summer, and Mr. Larry Huttshell was instrumental in translating XTRAN3S code.

INTRODUCTION

Aircraft structural design restrictions include strength criteria and aeroelastic instabilities. The design should ensure a satisfactory strength and minimum weight while maintaining a maximum operational flight velocity less than the critical aeroelastic instability velocity. Each aircraft also has a predetermined flight envelop. Figure 1 shows the flight range for various dynamic pressures.

Flutter is a phenomenon resulting from the unsteady interaction of the aerodynamic, structural, and inertial loads. For instance, large engines or stores on a wing can create a flutter problem. Also, a supercritical airfoil can have the effect of lowering the transonic flutter speed.

A recently developed USAF sponsored computer code called ASTROS is a multidisciplinary automated structural design tool. The code uses subsonic and supersonic steady and unsteady aerodynamics on a finite element model to generate a minimum weight structure, subject to stress and flutter constraints.

Although the aeroelastic instabilities are typically critical in the transonic flight regime, until recently, a non-linear aeroelastic analysis has been unattainable at the preliminary stages of design. The recent development of non-linear Computational Fluid Dynamic (CFD) computer codes makes it possible to perform transonic flutter analysis at initial design stages. CFD codes solve a set of non-linear

partial differential equations subject to a set of boundary conditions by discretizing the flow field and solving the resulting difference equations. XTRAN3S is a three-dimensional non-linear transonic small disturbance code which is capable of producing a transonic flutter boundary.

OBJECTIVE

The purpose of this investigation is to demonstrate the use of an unsteady, non-linear aerodynamic model as a tool in the preliminary structural design of an aircraft. By combining the structural optimization expertise of ASTROS with the transonic aeroelastic analysis ability of XTRAN3S, a transonic flutter boundary is established. Once the critical flutter point is determined, it is then possible to design a control law to be used in an active flutter suppression control system.

METHOD FOR STRUCTURAL AND AEROELASTIC DESIGN

A finite element model was created to represent the vehicle and served as input for ASTROS. The program was then instructed to compute the minimum weight structure for the vehicle to withstand a 9.0-g symmetric pull-up. ASTROS identified six anti-symmetric natural modes of vibration and performed a flutter analysis on the vehicle.

The flutter speed is determined by generating solutions at several dynamic pressures and by plotting the velocity versus the structural damping coefficient for each generalized coordinate. A damping coefficient of zero indicates a flutter occurrence for each dynamic pressure.

The purpose of the following steps is to extract the aerodynamic transfer function matrix which relates the degree of freedom to the corresponding loads. First, it is necessary to assume weak non-linearities and formulate a linear aerodynamic model for motion about the non-linear mean flow. The motion of the wing is constrained to one generalized coordinate, $q(t)g_i(x,y)$, where q is the pulse applied to the vehicle and g_i is the modal displacement. The wing is then perturbed and the resultant pressure time history, $F_i(x,y,t)$, is calculated. By integrating over the area of the wing, the generalized force is calculated.

$$Q_{ij}(t) = \int_A g_j(x, y) P_i(x, y, t) dA$$

The time history plots for the generalized coefficient and force are shown in Figures 2 and 3. The curves can be approximated by the sum of an appropriate combination of exponentials weighted by some proportion of the input function. The hatted variables indicate approximate functions.

$$\hat{Q}_{ij}(t) = (c_1 e^{b_1 t} + c_2 e^{b_2 t}) \cdot \hat{Q}(t)$$

These general equations are curvefit to the actual output, and the Laplace transform formula is applied to the approximate solutions.

$$\hat{Q}_{ij}(S) = \int_0^{\infty} \hat{Q}_{ij}(t) e^{-st} dt$$

$$\hat{Q}_i(S) = \int_0^{\infty} \hat{Q}_i(t) e^{-st} dt$$

Dropping the input from the generalized force equation produces the aerodynamic transfer function matrix.

$$A_{ij}(S) = \frac{\hat{Q}_{ij}(S)}{\hat{Q}_j(S)}$$

The aerodynamic transfer function can now be used to calculate the response created by an arbitrary vehicle motion.

$$\{ Q_i(S) \} = [A_{ij}(S)] \{ q_j(S) \}$$

Letting $S=ik$ and selecting a set of reduced frequencies (k) allows the aerodynamic transfer function to be used as input for an ASTROS flutter analysis. A new flutter boundary is established for the vehicle with the control system activated.

RESULTS

The model used in this investigation is a generic fighter aircraft. Its wing has ten spars and three ribs made from aluminum. The wing is characterized by a forty degree sweep, no twist, and a constant high speed airfoil. Assuming a fuselage weight of 16,000 lbs., ASTROS computed the minimum weight structure for the vehicle to withstand a 9.0-g symmetric pull-up. The optimization procedure assembled a full span wing box weight of 502 lbs, half of which is pictured in Figure 4. To induce a flutter situation, a 2250 lb store was attached to the rib near the midspan, and a 400 lb store was fixed on the next rib outboard.

ASTROS identified six anti-symmetric natural modes of vibration. The modes represent the following types of motion: rigid roll (Figure 5), first torsion (Figure 6), first and second bending, and the last two deform outboard of the center rib. A flutter analysis at Mach 0.85 was performed using subsonic doublet lattice theory and the PK flutter identification method. The flutter velocity was found to be 1136 ft/s at sea level, twenty percent greater than the actual speed of the aircraft. The flutter speed is within the 15% boundary displayed on the graph in Figure 7. The flutter frequency is 4.53 hz, and the primary contributors to the motion are the first two modes, rigid roll and wing torsion.

An unsteady aerodynamic model was constructed for XTRAN3S. The aerodynamic mode shapes were determined by

ASTROS and used as input for XTRAN3S. Due to the difference in the two aerodynamic grids, a FORTRAN interpolation program was written to determine the modal displacements on the XTRAN3S grid.

It was desired that the wing be perturbed with a ramp motion of the following form.

$$q(t) = \begin{cases} 0 & t \leq 0 \\ -\frac{A}{2} \left[1 - \cos\left(\frac{\pi}{a}t\right) \right] & 0 < t < a \\ -A & t \geq a \end{cases}$$

The Laplace transform of the input equation is the subsequent equation.

$$q(S) = -\frac{A}{2} \left[\frac{1}{S} - \frac{S}{S^2 + \left(\frac{\pi}{a}\right)^2} \right] [1 + e^{-aS}]$$

This function is not an input option for the XTRAN3S code, so it was necessary to modify the program to accommodate the above motion. However, without explicit knowledge of the code or access to its author, it was impossible to make the required changes. The generalized force is obtained by scaling the assumed transfer function form by the input.

$$\hat{Q} = -\frac{A}{2} \left[\frac{1}{S} - \frac{S}{S^2 + (\frac{\pi}{a})^2} \right] [1 + e^{-as}] \left[\frac{C_1}{S+b_1} + \frac{C_2}{S+b_2} \right]$$

Had the attempt at providing XTRAN3S with the desired input been successful, the output would have been curvefit to the following function, the inverse Laplace form of the above equation.

$$\hat{Q}(t) = -\frac{A}{2} \sum_{i=1}^n \left(\frac{C_i}{b_i} [u(t) - e^{-b_i t}] \right. \\ \left. - \frac{C_i}{(\frac{\pi}{a})^2 + b_i^2} \left[-b_i e^{-b_i t} + \frac{\pi}{a} \sin(\frac{\pi}{a} t) + b_i \cos(\frac{\pi}{a} t) \right] \right)$$

for $t < a$

$$\hat{Q}(t) = -\frac{A}{2} \sum_{i=1}^n \left(\frac{C_i}{b_i} [u(t) - e^{-b_i t}] \right. \\ \left. - \frac{C_i}{(\frac{\pi}{a})^2 + b_i^2} \left[-b_i e^{-b_i t} + \frac{\pi}{a} \sin(\frac{\pi}{a} t) + b_i \cos(\frac{\pi}{a} t) \right] \right) \\ - \frac{A}{2} u(t-a) \sum_{i=1}^n \left(\frac{C_i}{b_i} [u(t-a) - e^{-b_i(t-a)}] \right. \\ \left. - \frac{C_i}{(\frac{\pi}{a})^2 + b_i^2} \left[-b_i e^{-b_i(t-a)} + \frac{\pi}{a} \sin(\frac{\pi}{a}(t-a)) + b_i \cos(\frac{\pi}{a}(t-a)) \right] \right)$$

for $t > a$

The current approach is to use the ENSAERO Euler/Navier-Stokes code. An attempt at generating a transonic flutter condition is being conducted.

CONCLUSIONS AND RECOMMENDATIONS

The intent of this investigation is to establish a method for incorporating unsteady, non-linear, aerodynamics into the flutter free design of an aeroelastic model. The affordability of non-linear CFD computer programs makes it possible to perform essential transonic analysis. A flutter boundary can be determined for the complete range of Mach numbers during the preliminary design phase of the vehicle.

The high motivation to reduce computational cost makes a linear small disturbance code the desirable choice for establishing the flutter boundary. However, while the XTRAN3S small disturbance code is least expensive, the accuracy is limited due to the neglect of viscous effects and energy dissipation. ENSAERO is an Euler/Navier-Stokes solver which may generate a more reliable flutter analysis. This approach is now being pursued.

REFERENCES

Borland, C.J. and W.A. Sotomayer, "An Efficient Algorithm for Unsteady Transonic Flow", AIAA 7th Applied Aerodynamics Conference, AIAA 89-2230, JUL 1989.

Guruswamy, G.P., "Effect of Vorticial Flows on Aeroelasticity of Wings", AIAA 30th Structures, Structural Dynamics, and Materials Conference, AIAA 89-1183, APR 1989.

Neill, D.J., E.H. Johnson and R. Canfield, "ASTROS - A Multidisciplinary Automated Structural Design Tool", J. Aircraft, Vol. 27, No. 12, DEC 1990.

Flight Envelope

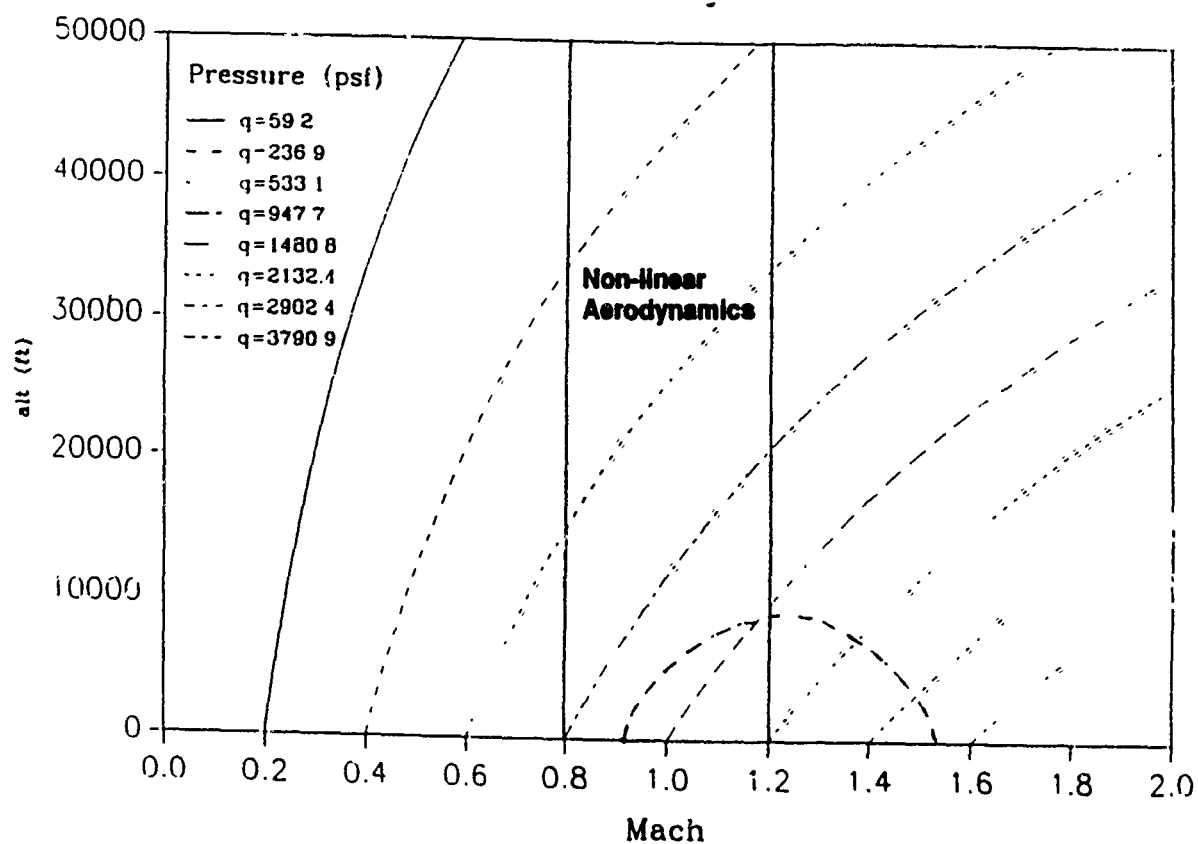


Figure 1. Vehicle Flight Envelope

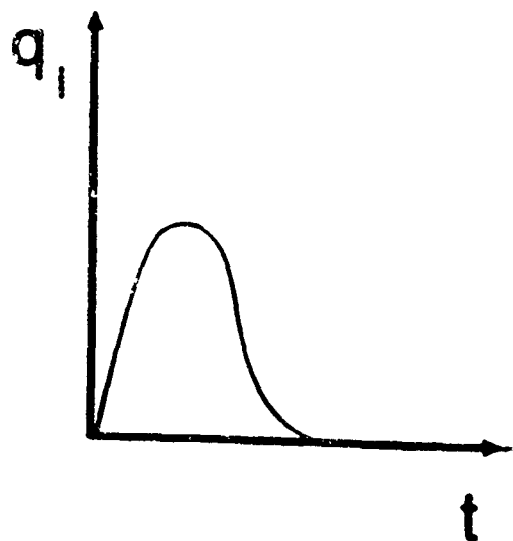


Figure 2. Generalized Coordinate Input Curve

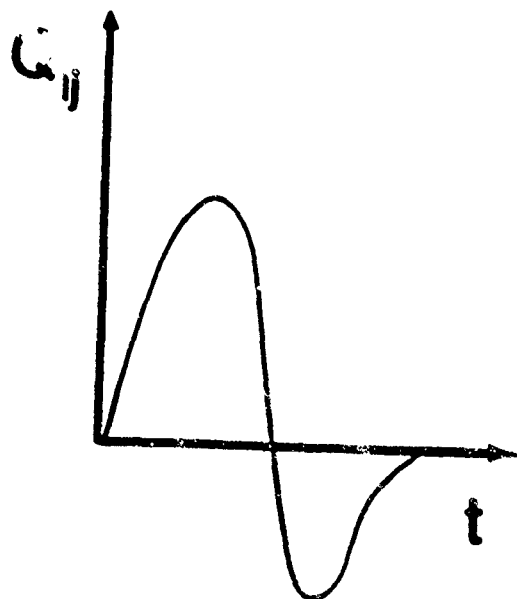


Figure 3. Generalized Force Output Curve

STRUCTURAL MEMBERS

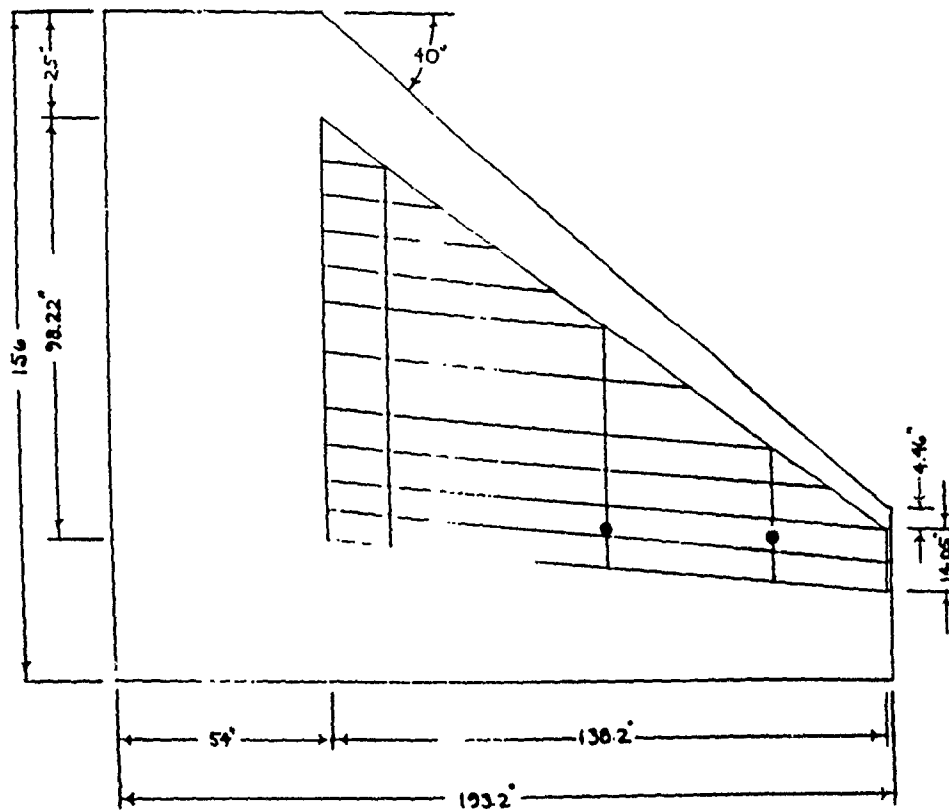


Figure 4. Wing Box Diagram

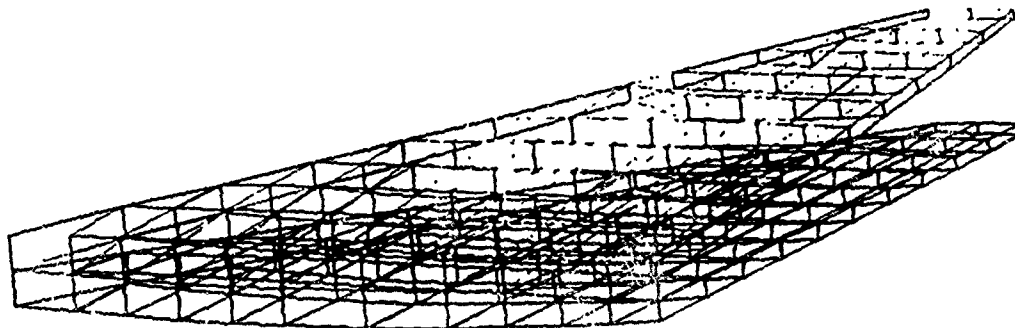


Figure 5. Rigid Roll Mode

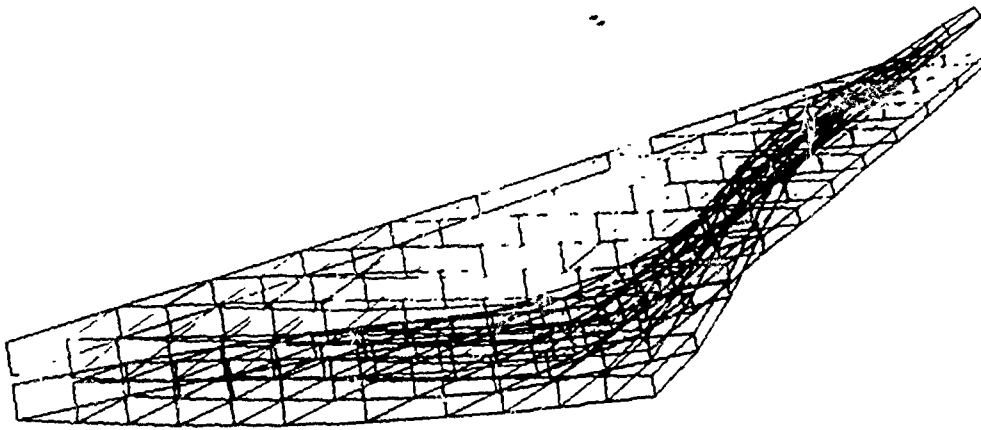


Figure 6. Torsion Mode of Vibration

FLUTTER MARGINS

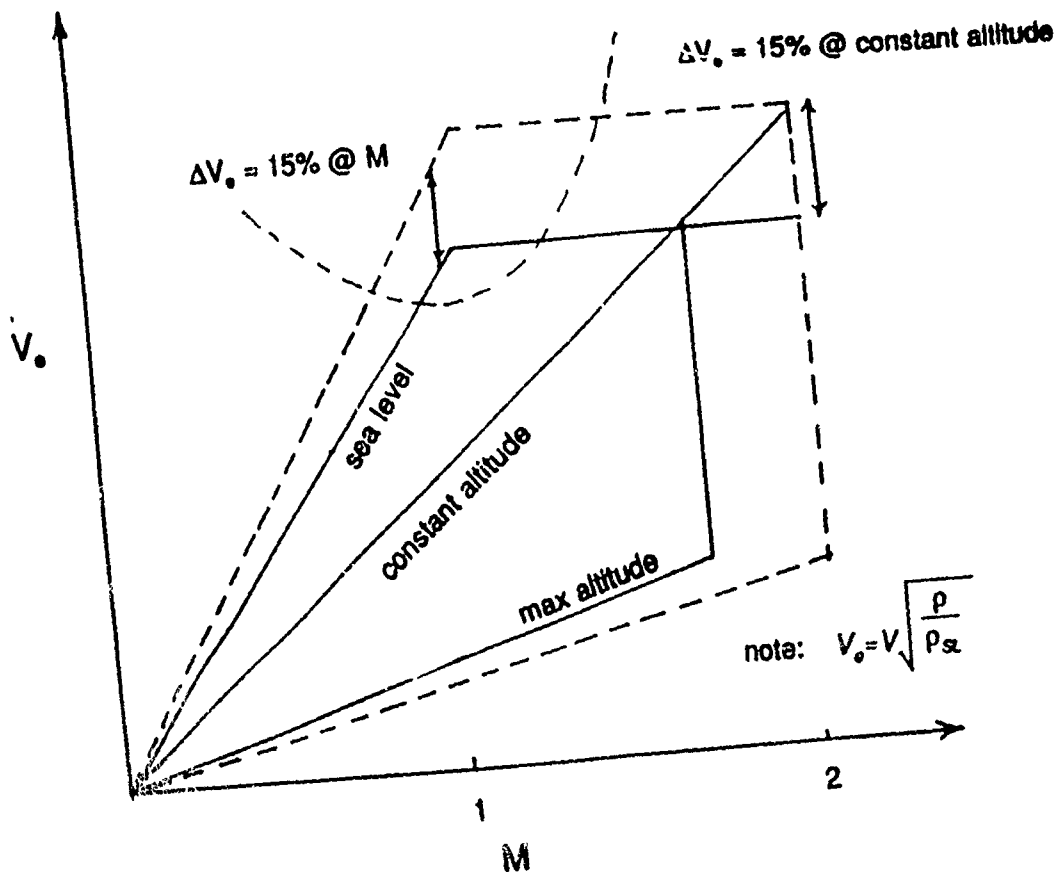


Figure 7. Flutter Margin with 15% Flutter Boundary

BALLISTIC IMPACT TEST OF GRAPHITE/EPOXY (COMPOSITE MATERIAL)

INSTRUCTOR AND PRINCIPAL INVESTIGATOR : DR. PIYUSH DUTTA

ARMY CORPS OF ENGINEERS
COLD REGIONS RESEARCH & ENGINEERING LABORATORY
72 LYME ROAD
HANOVER, NH 03755-1290

PERFORMED AT U.S. ARMY COLD REGIONS RESEARCH AND ENGINEERING LABORATORY
(USACRREL)

PERFORMED, PREPARED AND WRITTEN

BY: AZIZ TADAYON

SUMMER 1991

ACKNOWLEDGEMENT

I wish to thank Dr. David Hui, my thesis advisor, for his constant support and brainstorming discussions. I would also like to thank the Air Force System Command, Air Force Office of Science Research for their initial sponsorship of this research.

I wish to thank USA CRREL for continued sponsorship of this research, excellent facilities and pleasant working environment.

I would also like to thank Dr. Piyush Dutta for his critical help, advice and guidance. Special thanks goes to Dennis Farrell and John Kalafut for their exceptional help in acquiring supplies and solving technical problems, and to Susan Taylor for helping me with SEM photos and to Pam Bosworth for typing the results of this scientific experiment.

ABSTRACT

This paper is explaining the experimental work of 1/2 in. diameter steel ball impacting a graphite/epoxy panel (composite materials) with two different velocities:

a.) low velocity (400 ft/sec), is done under atmospheric pressure 14.7 psi or 760 Hg.

b.) high velocity (3200 ft/sec), is done under vacuum conditions to eliminate the air resistance factor.

Two different types of powder were used:

1. unique, which is ideal for low velocity.
2. red dot which burns much better under vacuum conditions and is suitable for high velocity.

The study and analysis of fragments (image processing size and shape distribution) under SEM and C-Scan machines are the main goals of this scientific phenomena.

INTRODUCTION

The graphite/epoxy panels (composite materials) that are used as targets in this experiment were made of materials having properties of several distinct components. These panels are comprised of several thin layers that lay on each other smoothly with different orientation (0° , 45° , 90° ...). The thickness of each layer is nearly 0.5 mm.

Two types of velocities were obtained in this experiment:

- a. Low velocity (400 ft/sec).
- b. High velocity (3200 ft/sec).

The low velocity, which is also called V_{50} ballistic impact, which is defined as the velocity at which a specified round has a 50% probability of perforating the target (graphite/epoxy panel) when striking it at 400 ft/sec (low velocity). This was the case at low velocity which were not perforate and penetrate the target completely.

For the high velocity (3200 ft/sec) which was done under vacuum condition in order to eliminate the air resistance, the target (graphite/epoxy panel) was perforated and penetrated completely and cleanly because of very high speed.

The experimental values have been compared with theoretical prediction and also the experimental results vindicate the fracture mechanic theory.

OBJECTIVE AND GOALS

The main purposes of this experiment are as following:

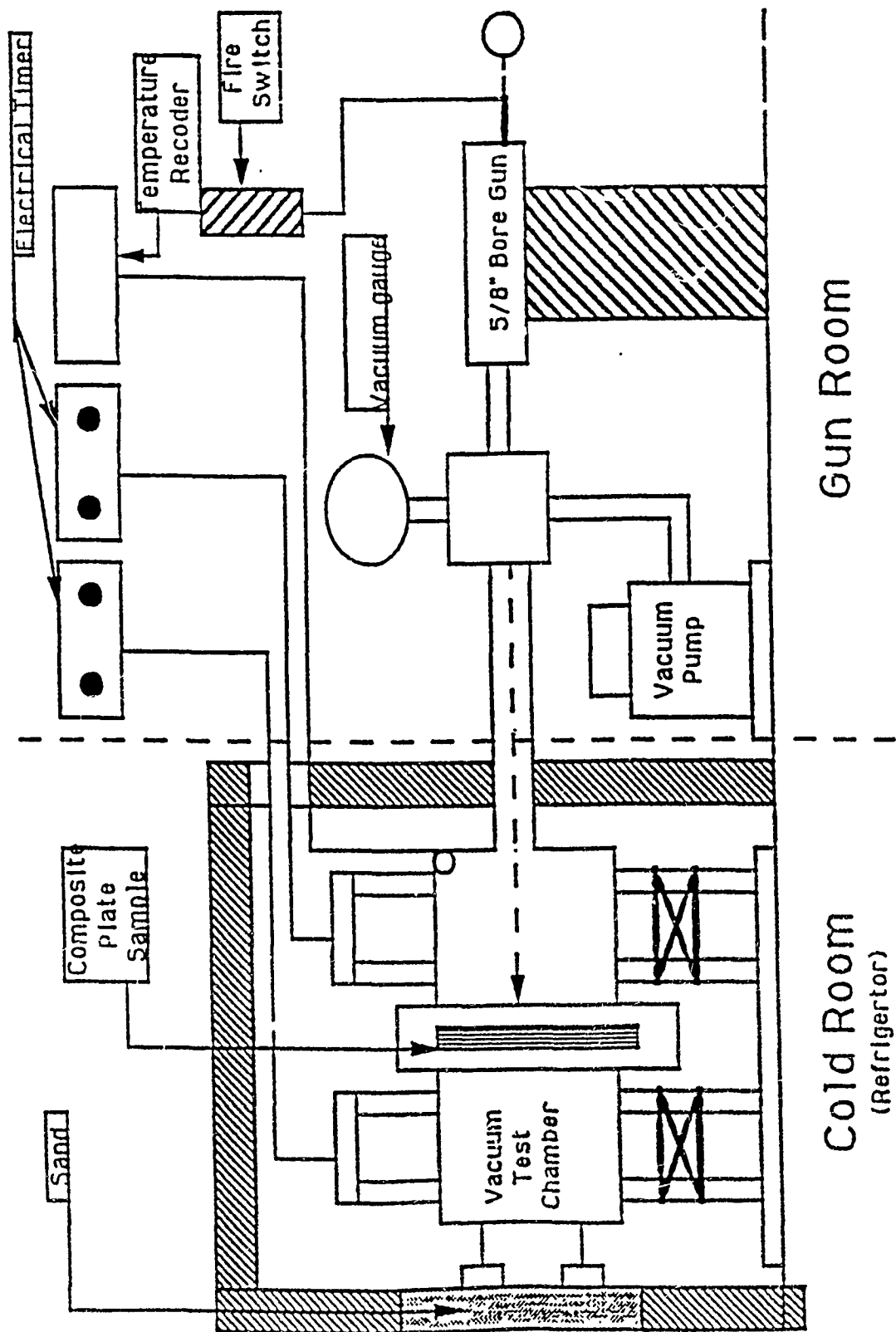
1. To obtain low velocity (400 ft/sec).
2. To obtain high velocity (3200 ft/sec).
3. Study, analyze and process images of fragments, size and shape distribution under SEM (scanning electron microscopy).
4. Using C-SCAN to determine the volume and area damaged during these impacts.
5. To determine and calculate the high energy absorption of graphite/epoxy panels.
6. To compare the results and values of the experiment with theory of Fracture mechanic and its predictions.

PROCEDURE

To give you a taste of this scientific engineering experiment, I would like to refer you to the schematic set up of this task shown on next page.

As you can see, the whole set up of this project consists of two main parts as follows:

1. Gun Room [firing room].
 2. Cold room [refrigerator].
-
1. Gun Room, comprised of:
 - a. 5/8" steel bore gun whose length is 23".
 - b. An aluminum pipe connects the bore gun to the first plexiglas (situated in cold room).
 - c. A vacuum pump and vacuum gauge which are used for high velocity (3200 ft/sec) and located between bore gun (in gun room) and first plexiglas (in cold room).
 2. Cold Room (Refrigerator) consists of the following items:
 - a. Two (2) plexiglas with the diameter of 12" and length of 18". Each plexiglas has a ring with 21" in diameter that the composite plate sample (graphite epoxy panel) is sandwiched by them.
 - b. Two chronograph light screens covering each plexiglas from outside and they are connected to the electrical timers (time recorders) for measuring the time.
 - c. A flypaper of 36" x 18" with double coated sticky tape placed inside each plexiglas to capture the fragments.
 - d. A sandbag is the final destination of the projectile. The sandbag has been inserted in the front wall of the cold room.



SCHEMATIC BALLISTICS IMPACT TEST

VELOCITY CALIBRATIONS

After many attempts with two different powders (unique and red dot), I concluded that unique powder is good for low velocities while red dot powder burns better at high velocities.

Calibration of velocities (low and high) are as follows:

- a. LOW VELOCITY (400 FT/SEC) CALIBRATIONS: As I mentioned above, unique powder is used for these velocities. To find the trend and repeatability of 400 ft/sec velocities many attempts were made as you can see in the table below (Table 1).

Table 1. Low Velocity Calibration.

Unique				
Test #	* grains	T ₁ (μsec)	T ₂ (μsec)	T ₃ (μsec)
1	5.0	0.003178	0.003203	0.003173
Velocity (ft/sec)		(314.7)	(312.2)	(315.1)
2	5.0	0.002287	0.002307	0.002285
Velocity (ft/sec)		(437.2)	(433.5)	(437.6)
3	5.0	0.00297	0.002621	0.002594
Velocity (ft/sec)		(385.0)	(381.5)	(385.5)
4	5.0	0.002927	0.002987	0.002982
Velocity (ft/sec)		(341.6)	(334.7)	(335.3)
5	5.0	0.002963	0.002988	0.002958
Velocity (ft/sec)		(337.5)	(334.7)	(338.1)

* 1 lb = 7000 grains, 1 gram = 15.43 grains

Since our objective is to obtain 400 ft/sec velocity, I increased the amount of unique powder as it shown the trials in below table; (Table 2).

Table 2. Low Velocity Calibration (With average velocities).

Test #	Uniqu * grains	T ₁ (μsec)	T ₂ (μsec)	T ₃ (μsec)	Average velocity
1	5.2	0.003554	0.003639	0.003624	277.3
Velocity (ft/sec)		(281)	(275)	(276)	
2	5.2	0.003252	0.003305	0.003271	305.2
Velocity (ft/sec)		(307.5)	(302.6)	(305.1)	
3	5.2	0.003469	0.003446	0.003458	289.3
Velocity (ft/sec)		(288.3)	(290.2)	(289.5)	
4	5.3	0.002748	0.002795	0.002792	360
Velocity (ft/sec)		(363.9)	(357.7)	(358.2)	
5	5.3	0.0024849	0.0023081	0.0022824	424.6
Velocity (ft/sec)		(402.4)	(433.2)	(438.1)	
6	5.3	0.0024047	0.002449	0.0024413	411.2
Velocity (ft/sec)		(415.8)	(408.3)	(409.6)	
7	5.3	0.002661	0.002657	0.002660	376.1
Velocity (ft/sec)		(376)	(376.4)	(375.9)	
8	5.3	0.0029998	0.0029898	0.0030480	332.0
Velocity (ft/sec)		(333.4)	(334.0)	(328.1)	
9	5.3	0.0023015	0.0022567	0.002345	434.6
Velocity (ft/sec)		(434.5)	(443.1)	(426.2)	
10	5.4	0.002318	0.002780	0.002275	435.4
Velocity (ft/sec)		(431.0)	(438.6)	(439.6)	
11	5.4	0.002134	0.002152	0.002128	467.8
Velocity (ft/sec)		(468.6)	(464.7)	(470.0)	

As you can see the trend and repeatability of velocities and their closeness values.

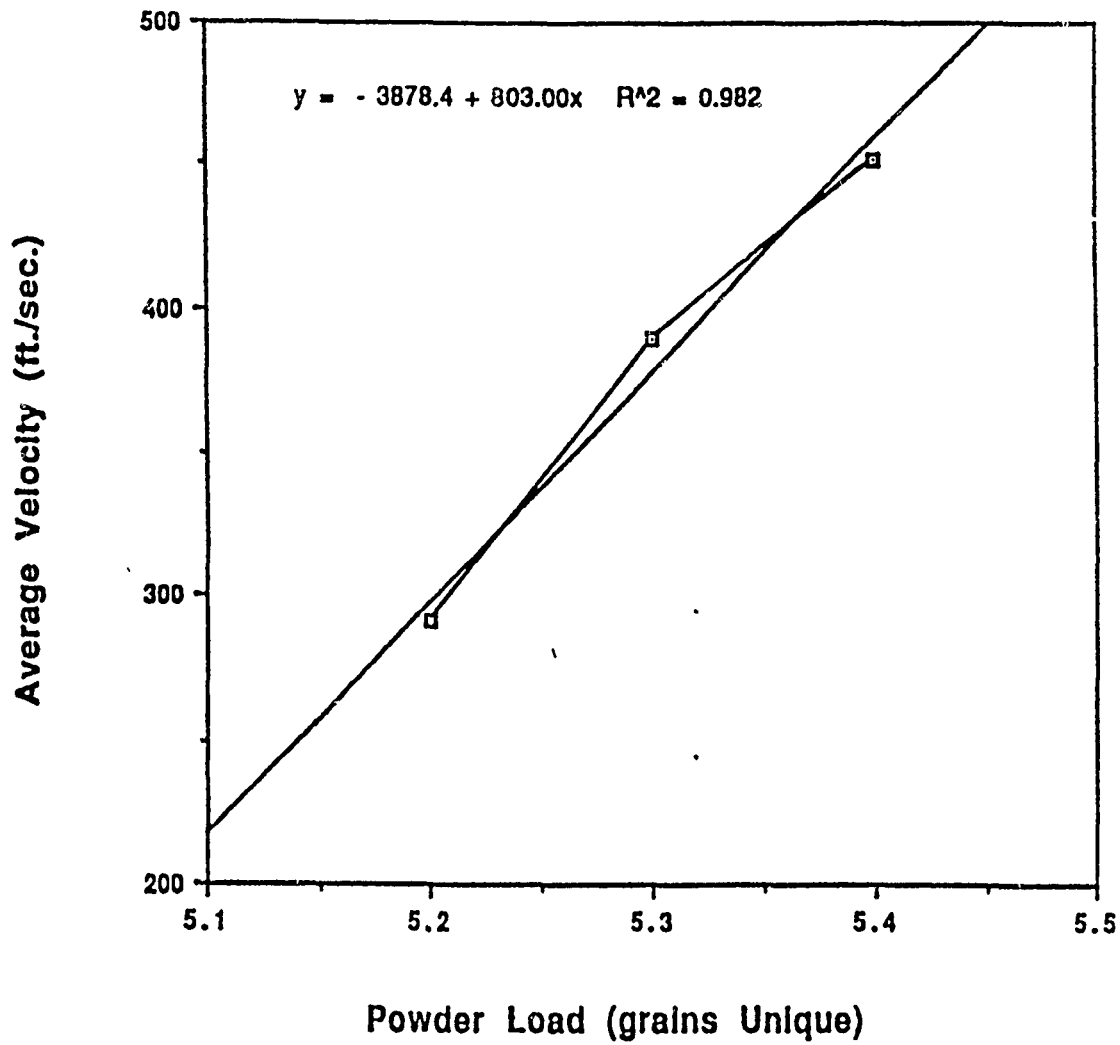
To be more accurate and precise I refer you to the below table (Table 3)

Table 3. Low Velocity Calibration (with their second average velocities).

Test #	Grains (unique)	Average Velocity (1) ft/sec	Average velocity (2)
1	5.2	277.3	291 ft/sec
2	5.2	305.2	
3	5.2	289.3	
4	5.3	360.0	390.0 ft/sec
5	5.3	424.6	
6	5.3	411.2	
7	5.3	376.2	
8	5.3	332.0	
9	5.3	434.6	
10	5.4	435.4	451.6 ft/sec
11	5.4	467.8	

With these values (# of grains vs average velocity 2) I constructed the following graph that we can see a linear relationship between powder load and velocities [graph 1].

POWDER LOAD VS. VELOCITY



STUDY, ANALYSES AND IMAGE PROCESSING OF FRAGMENTS UNDER SEM MACHINE

First some background of SEM:

The scanning electron microscope (SEM) is one of the most versatile instruments available for the examination and analysis of the microstructural characteristics of solid objects. The primary reason for the SEM's usefulness is high resolution which can be obtained when bulk objects are examined, values of the order of 5 nm (50 Å) are usually quoted for commercial instruments.

Another important feature of the SEM is the three-dimensional appearance of the specimen image, which is a direct result of the large depth of field. The SEM machine is also capable of examining objects at very low magnification.

Study of Fragments of low velocity under SEM:

Four SEM mounts were made of shrapnel from multi-ply graphite epoxy composites; two shots were fired at low velocity and particles were picked from the front and back collection areas (plexi 1 and 2) of both shots (panel 1 and 2). Four particles from each collection area, in one case five (panel 1 plexi 2), were mounted for observation. I tried to select particles which represented the different types generated. Most particles appeared to be needle-shaped fragments. There were also some rounded particles and some large pieces which had graphite fibers protruding at a 45° and 90° angle from one another. As with other studies of graphite epoxy composites high magnification X800 photos were made of fiber surfaces to: estimate and compare the number of heckle marks; look for cracks and make

note of any features which indicate the failure mechanism. Care needs to be taken to mount the particles exactly as they were on the fly paper so that the adhesive is not imaged.

Figure 1 shows the collective particles selected from panel 1 plexiglas (front side), they are all together four particles (long needle, small needle, large round and medium round) as Figure 1 shows, before each particle individually analyzed under SEM machine.

The four particles of panel 1 plexiglas 1 individually put under SEM for observation and analysis, the following observation were concluded:

Panel 1, Plexi 1: The mount has two needle shaped particles, one several mm long and the other a little over a mm. They were photographed at a magnification of x800 for comparison with other fragments and with microphotographs of surfaces of graphite epoxy composites failed under tension and compression. The surfaces of these two fragments show fairly clean graphite fibers imbedded in the epoxy. The small needle has some heckle marks where fibers have been removed. There is also an epoxy-rich area with no heckle marks and a smeared out appearance to the epoxy. The large needle has few heckle marks and a lot of epoxy fragments on its surface.

In addition to the two needle-shaped particles, there is also a round disk-shaped particle and an irregularly shaped particle which resembles a tear drop. Neither of these particles had any fibers visible on their surfaces. The shape of the fragments and the presence of bubbles on their surfaces, suggest that they were melted and are epoxy rich.

Figures 2 and 3 show the particles of panel 1 Plexiglas 1 under SEM machine and their closeup.

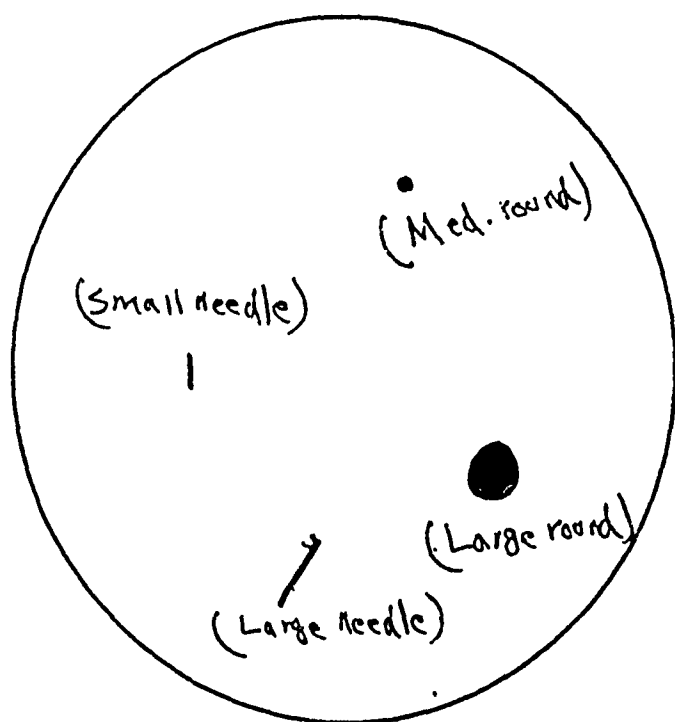


FIGURE 1: 4 Particles from Panel 1, Plexiglass1
(large & small needles, large & medium round
particles)

Large
Round



Panel 1 Plexi 1

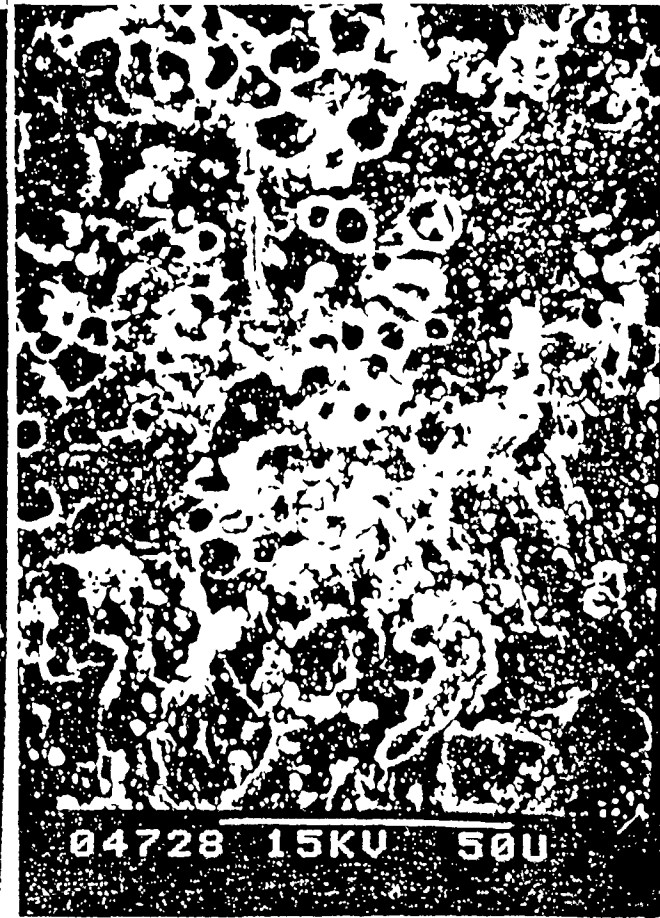


Panel 1 Plexi 1

Med.

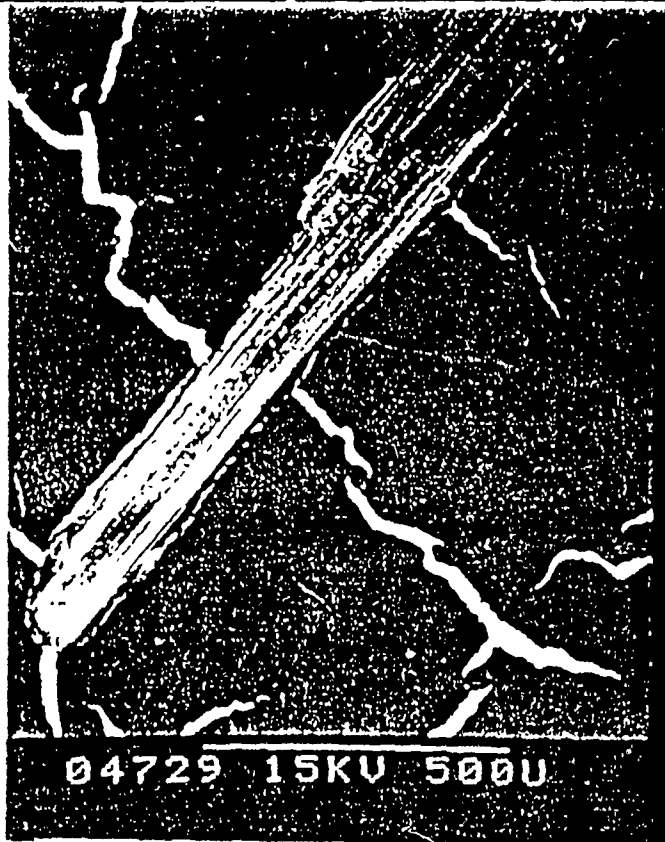


Panel 1 Plexi 1



Panel 1 Plexi 1 x 800

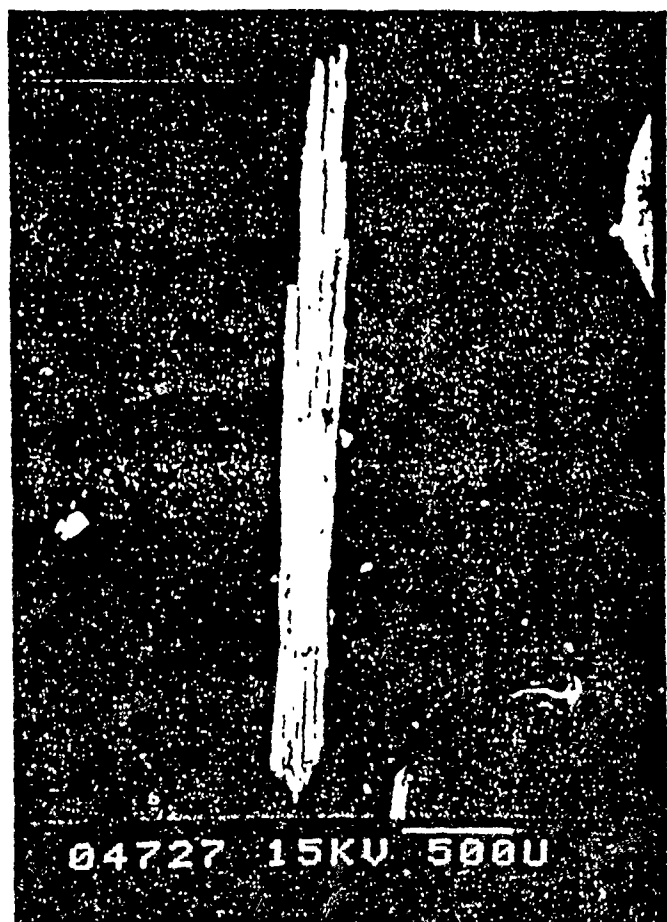
FIGURE 2: LARGE & MEDIUM PARTICLES OF PANEL 1, PLEXI 1 UNDER SEM & THEIR
800 MAGNIFICATION.



Panel 1 Plexi 1 Small needle



Panel 1 Plexi 1 Small needle



Panel 1 Plexi 1 Large needle



Panel 1 Plexi 1 Large needle x 800

FIGURE 3: LARGE & SMALL NEEDLE PARTICLES OF PANEL 1 PLEXI 1 UNDER SEM & THEIR 800 MAGNIFICATION.

For Panel 1 plexiglas 2 of low velocity five particle were picked up (large, medium, small particles and large and small needle) for SEM observation and analysis. Figure 4 shows these particles collectively and before individually being studied under SEM machine.

The five fragments of panel 1 plexiglass 2 individually put under SEM machine for analyses and observation, the following results were observed.

Panel 1, Plexi 2: The mount has two needle shaped particles, a half moon shaped particle which appeared to be half of a round disk and two multi-ply particles. The half moon particle has no visible graphite fibers and its surface has a smeared look. The two multi-ply particles each have two layers of fibers at 45° from each other. In the case of the large multi-ply particle it has had large fiber pull-out sections running parallel to the fiber direction; removed when the ply, which used to be on top and has fibers at 90° , was detached. There is a layer of epoxy showing the orientation of the fibers and also showing well developed heckle marks. The grooves where the fibers used to be are clean and smooth indicating that most of the energy went into breaking the epoxy. The smaller multi-ply particle shows some cracking along the fiber length and heckle marks. The latter are smaller than those seen in the larger multi-ply particle. The large needle has many fibers but few grooves where the fibers have been removed. It resembles the small multiply particle in this regard and also in the size of heckle marks. The small needle has mainly grooves where the fibers have been removed and only a few areas with heckles.

Figure 5, 6, 7 show the five fragments of graphite/epoxy of panel 1 plexiglas 2 with their 800 magnifications.

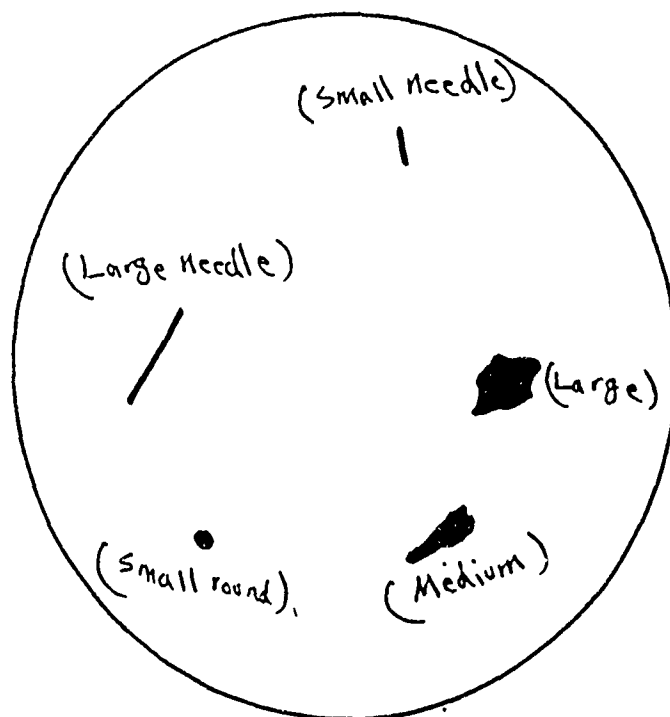


FIGURE 4: Panel 1, Plexiglass 2, 5 Fragments of graphite epoxy (large, medium, small particles, large and small needles).

Large
Fragment



Panel 1 Plexi 2 x 30 Large multiply particle

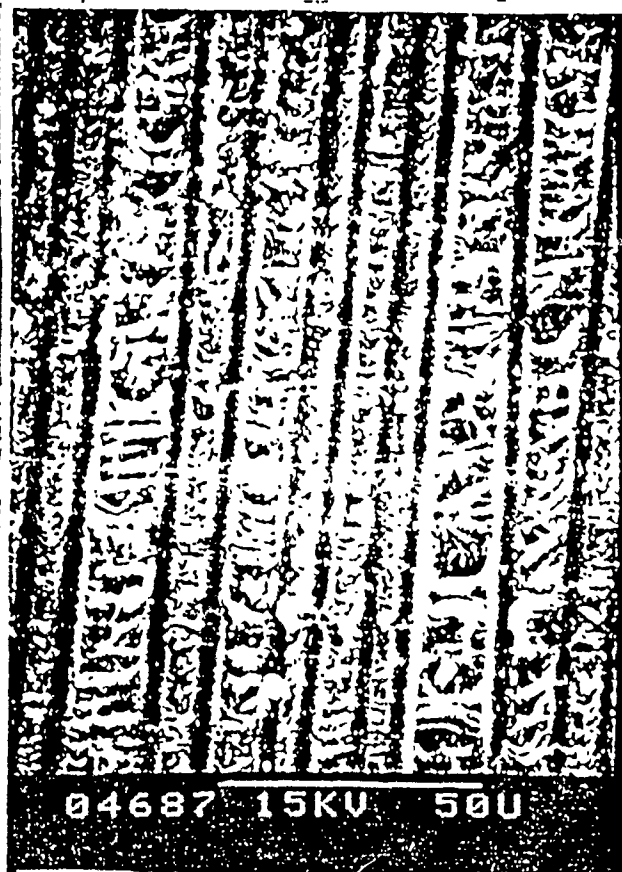


Panel 1 Plexi 2 Large multiply particle

Med.
Fragment



Panel 1 Plexi 2 Medlum



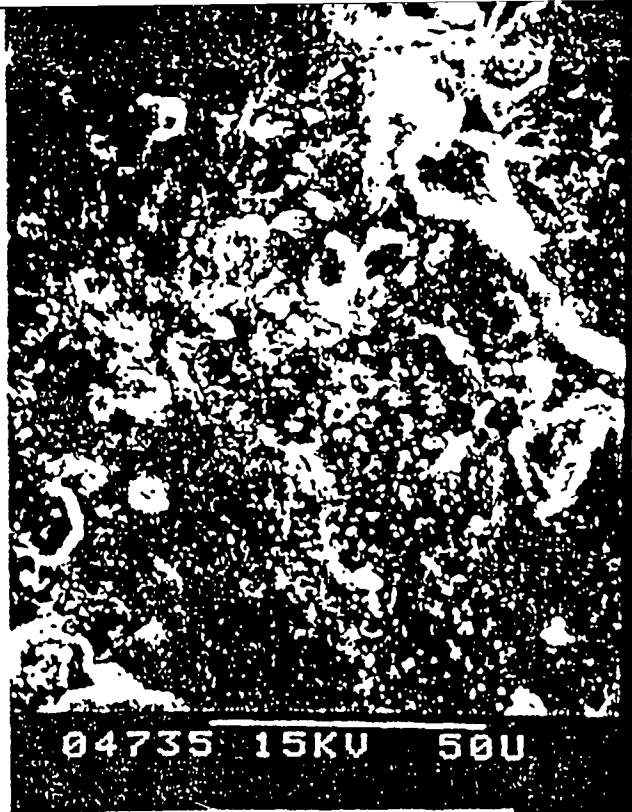
Panel 1 Plexi 2 Multiply (med) particle

FIGURE 5

Small
Round
(Rear)

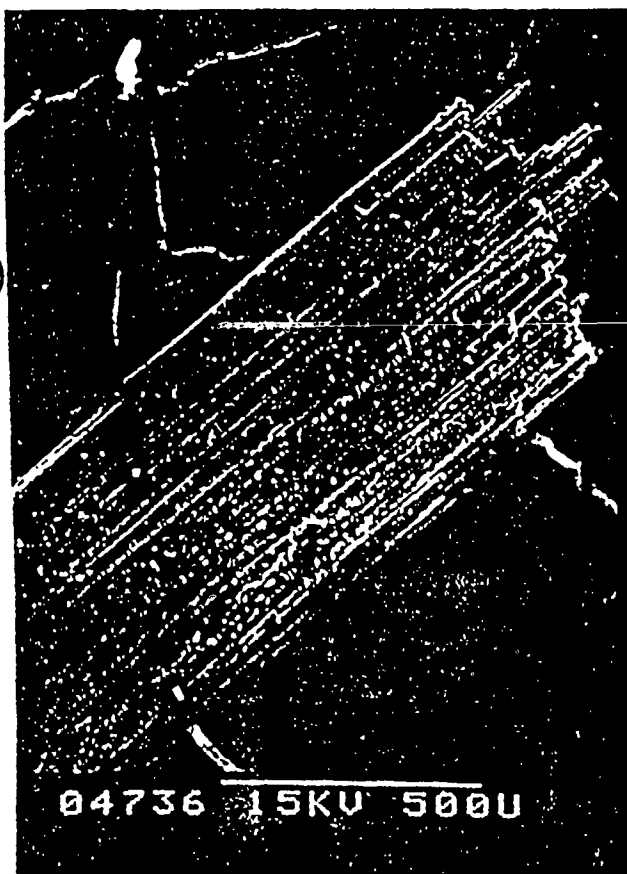


Panel 1 Plexi 2 Half moon particle



Panel 1 Plexi 2 x 800 Half moon
particle

(Large
Needle)

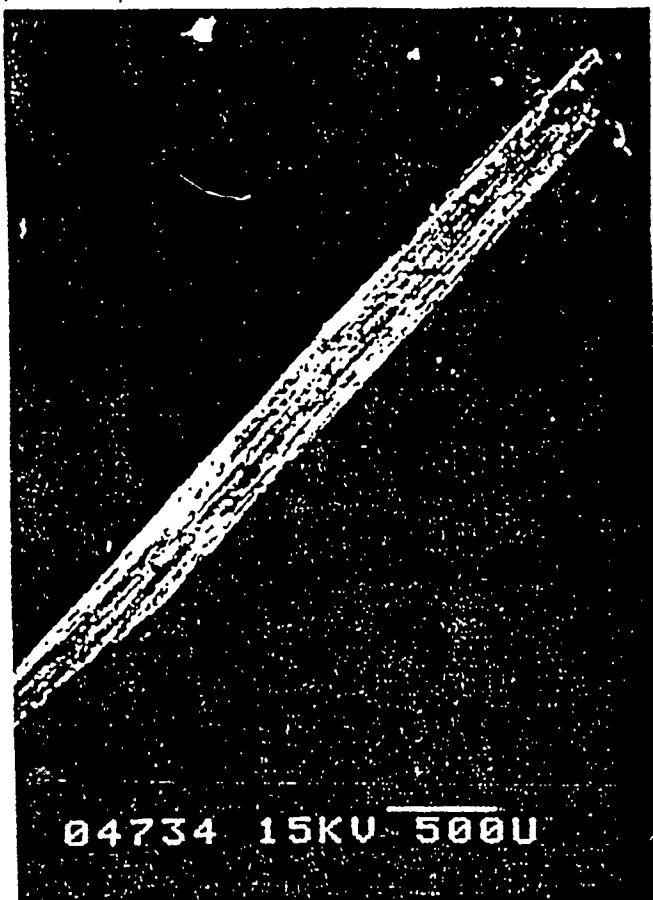


Panel 1 Plexi 2 Large needle



Panel 1 Plexi 2 Large needle

FIGURE 6



Panel 1 Plexi 2 x 30 Small needle



Panel 1 Plexi 2 Small needle

Figure 7: Small needle fragment of Panel 1, Plexi 2 and its 800 magnification.

b. HIGH VELOCITY (3200 FT/SEC) CALIBRATIONS:

I mentioned earlier that red dot powder burns better at high speeds and under vacuum conditions. Therefore it was used to calibrate the high velocity under vacuum.

Calibrations started with 10 grains of red dot powder under constant vacuum of 800 millitorr as shown in below table (Table 4).

Table 4: Calibration for high velocity (3200 ft/sec).

	Grains	Vacuum	Velocity
<u>Test #</u>	<u>(Red dot)</u>	<u>(millitorr)*</u>	<u>ft/sec</u>
1	10	800	2732
2	10	800	3185
3	10	800	2439
4	10	800	1912

The average of these velocities is 2567 ft/sec.

$$* 1 \text{ torr} = 133.3 \text{ pascal} \rightarrow \text{Pascal} = \frac{1}{133.3} = 0.00751 \text{ Torr}$$

$$1 \text{ PSI} = 6894 \text{ Pascal} \rightarrow \text{Pascal} = \frac{1}{6894} = 0.000145 \text{ PSI}$$

$$0.00751 \text{ Torr} = 0.000145 \text{ PSI}$$

$$\text{Torr} = \frac{0.000145}{0.00751} \text{ PSI} = 0.0193 \text{ PSI}$$

$$1 \text{ Torr} = 0.0193 \text{ PSI}$$

$$1 \text{ Millitorr} = \frac{0.0193}{1000} \text{ PSI} = 0.0000193 \text{ PSI}$$

in our case which is 800 millitorr will be

$$800 \times 0.0000193 \text{ PSI} = 0.01544 \text{ PSI}$$

$$800 \text{ millitorr} = 0.01544 \text{ PSI}$$

As you can see the relationship between grains, velocities and vacuum from Table 4. From above conversions and calculations, we can see that the 800 millitorr = 0.01544 PSI which is very small amount of air in the system, therefore I decided to increase the amount of powder to 13 grains in order to obtain our desirable heigh velocity (3200 ft/sec). While keeping the vacuum constant on 800 millitorr. [Vacuum pump did not work properly at most of the time, that's why I kept it constant at 800 millitorr].

The following data were obtained when the amount of grains were increased to 13 grains, and constant vacuum (800 millitorr).

Table 5: Calibration for high velocity (3200 ft/sec).

<hr/>			
	Grains	Velocity	Vacuum
<u>Test #</u>	<u>(red dot)</u>	<u>ft/sec</u>	<u>(millitorr)</u>
1	13	3183	800
2	13	3208	800
3	13	3227	800
<hr/>			

We can see the closeness of velocities with 13 grains of red dot powder under vacuum (800 millitorr). The average of these velocities is 3206 ft/sec which is our desirable speed for high velocity.

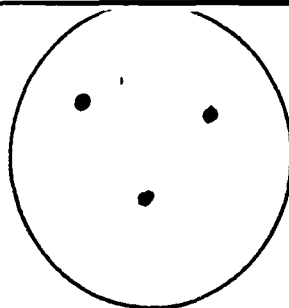
I fired a shot with this desirable velocity at graphite/epoxy panel, captured its fragments, and made them ready for SEM analysis.

STUDY OF FRAGMENTS OF HIGH VELOCITY UNDER SEM

Nine SEM mounts were made of shrapnel from multi-ply graphite epoxy of high velocity (3200 ft/sec). Fragments were picked from the front (plexi 1) and back or rear (plexi 2) of collection areas. Five mounts for plexiglas 1 and four mounts for plexiglas 2 were made. I tried to select similar fragments for each mount to see whether we can get similar results from similar particles. It is sort of a lengthy and tedious job, but it is necessary to find out. Figures 8 and 9 show these nine mounts of plexiglas 1 (5 mounts) and plexiglas 2 (4 mounts) individually before being examined under SEM machine.

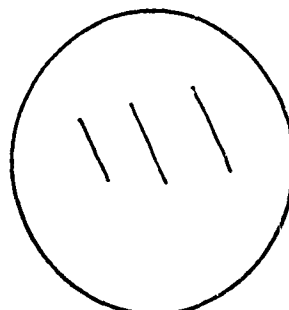
As I put these nine mounts of fragment in SEM machine individually and each fragments (particles) of each mounts individually being studied under SEM the following conclusions being observed; for high velocity:

Mount #1



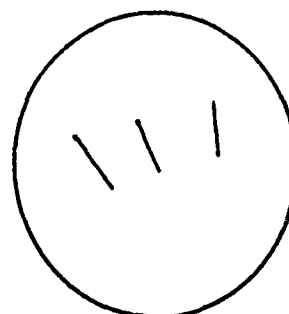
Three rounded Particles (fragments)

Mount #2



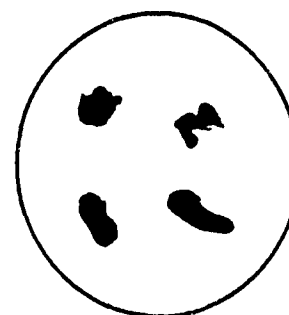
Three Large Needles

Mount #3



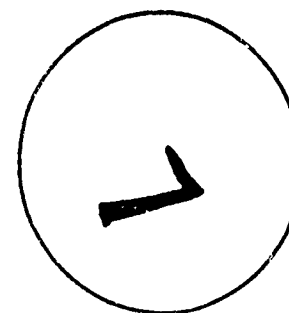
Three Small Needles

Mount #4



Four Irregular Particles

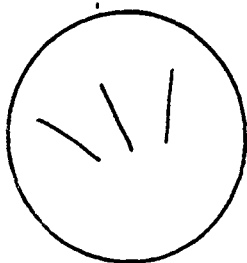
Mount #5



Single Multi-p' Particle

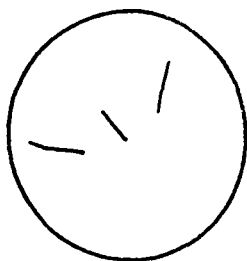
FIGURE 3: Five mounts of high velocity fragments (panel 1, plexi 1)

Mount #1



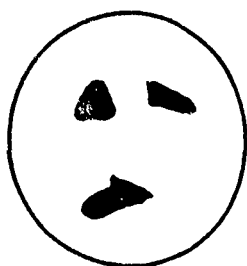
Three Large Needles

Mount #2



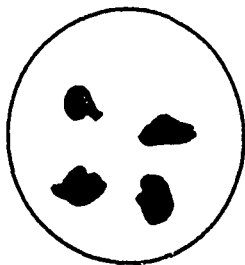
Three Small Needles

Mount #3



Three Multi-ply Particles

Mount #4



Four Irregular Particles

FIGURE 9: Four mounts of high velocity fragments
(Panel 1, Plexi 2)

Panel 1 Plexi 1:

In spite of low velocity fragments, high velocity fragments had fewer rounded particles than did the low velocity particles. Those were found only on the front side (plexi 1). None of the rounded particles had bubbled surfaces (see Figs. 10 and on for individual fragments under SEM and their closeup, magnification 800). They looked less smooth and had multi-layered surface.

Panel 1 Plexi 2:

It had many needle shaped particles both large and small. Some large scale - 10 cm pieces of composite and many multiply partocles.

The characteristics of the needle shaped and multi-ply particles do not seem different from those of the low velocity fragments, although not a bit of time was taken in comparing them.

Some surfaces are mainly expoxy and show the broken fragments of grooves left when the top layer of fibers were removed.

There are also very smooth surfaces where the fibers are present and appear to have debanded smoothly from the epoxy. Little epoxy adhered to their exposed surfaces.

Figures 10 and on show the individual fragments under SEM machine and their x 800 magnifications. These figures are for high velocity (3200 ft/sec) and from panel 1 plexi 1 and panel 1 plexi 2, as it have mentioned at the bottom of each one individually.

CONCLUSIONS

In this scientific engineering experiment the low velocity (400 ft/sec) and high velocity (3200 ft/sec) were obtained through trial/errors and repeatability. Both fragments for low and high velocities, front (plexi 1) and back (plexi 2) were studied individually under SEM machine. The last part of this experiment which is calculation of damaged volume, damaged area has to be done at Wright Patterson AFB in Ohio. For these calculation c-scan need to be used and that is available at Wright Patterson, AFB.

Subsequently, after these calculation, high energy absorbtion of graphite epoxy panel (composite materials) can be obtained which was the ultimate goal of this challenging research.

Also it is worthy to note that in high velocity fragments there were not too many heckle marks as we saw in low velocity fragments, because of very high speed at very short time.

**HYPERSONIC VEHICLE CONTROL STRATEGIES:
PRELIMINARY CONSIDERATIONS**

**Dr. Daniel J. Biezad
AFOSR 1991 Faculty Research Program**

**Mr. Phuong Vu
AFOSR 1991 Graduate Student Research Program**

**Aeronautical Engineering Department
Cal Poly State University
San Luis Obispo, CA 93407**

presented to

**Mr. Chuck Suchomel
AFOSR Research Focal Point
WL/FIGC
WPAFB, Ohio 45433-6553**

" This document may contain information subject to the International Traffic in Arms Regulation (ITAR) or the Export Administration Regulation (EAR) of 1979, which may not be exported, released, or disclosed to foreign national inside or outside the United States without first obtaining an export license. A violation of the ITAR or EAR may be subject to a penalty of up to 10 years imprisonment or fine of not more than \$1,000,000 or both under 22 U.S.C. 2778 or Section 2410 of the Export Administration Act of 1979. Include this notice with any reproduced part of this document. "

ABSTRACT

Preliminary research in the area of hypersonic vehicle (HSV) control was accomplished to prepare for a proposal submission to AFOSR (including an extensive literature review). Linearized state-space models of a generic HSV configuration, obtained from Systems Technology Inc (STI), were catalogued and investigated using the controls analysis packages "CC Version 4" for IBM-compatibles and using MATLAB/SIMULAB for the MAC computer family. Tutorial and analysis "command files" were generated in "MACRO" form (CC-IBM) and in "M-FILES" (MATLAB-MAC).

The following research goals were established and investigated as time permitted:

- (1) Identify the governing key control issues in hypersonic "equilibrium" flight.
- (2) Investigate viable longitudinal loop closure control strategies (called "minimal SCAS" strategies by STI) for a succession of equilibrium points in the hypersonic envelope that define a particular mission phase.
- (3) Monitor and prevent excessive control deflection and deflection rates.
- (4) Apply new MIMO stability and robustness metrics which are applicable to the control strategies and present them in a form suitable for analysis.
- (5) Identify and apply handling qualities metrics which will provide insight into pilot-in-the-loop hypersonic flight, especially in the vital area of predicting flying qualities "cliffs".

1. INTRODUCTION.

Control of hypersonic vehicles (HSV) presents a non-linear, time-varying problem of enormous complexity. At hypersonic speeds the "thin" shock layers and "thick" boundary layers cause viscous and chemical heat interactions that severely complicate the determination of forces and moments. This is further aggravated by the large, lifting bow wave under the vehicle, by the propulsive coupling inherent in HSV designs, and by the amplified effects of static margin changes with Mach number.¹

HSV equations of motions take into account a spherical earth, a rotating "local vertical" direction, and centrifugal relief.² These considerations complicate the traditional equations of motion for a flat, non-rotating earth and introduce new dynamic modes.³ Their impacts are primarily in the long period modes (phugoid, height, density).⁴ Considering the magnitude of the other difficulties, however, these new individual terms in HSV linearizations of the equations of motion may be of less concern than the very concept of linearization itself. Equilibrium is difficult to define for hypersonic flight, powered or unpowered, and typically corresponds to a "nominal" ascent or descent trajectory where forward acceleration (dV_T/dt) is a constant not equal to zero.⁵⁻⁷ It is assumed that the trim required for this type of operation will not be excessive or distracting from the pilot's point of view.

These considerations make it clear that any flight control designs based on linear models may expect considerable modification after "first cut" control laws are developed. These initial control laws may be referred to as minimal stability, control, and augmentation systems ("minimal SCAS"), and represent the augmented aircraft as seen by the pilot.^{8,9} Even though these laws are preliminary, however, they incorporate significant characteristics of hypersonic flight and will be valuable for the insight they provide as well as for their potential utility in design. For example, any "minimal SCAS" architecture has to take into account the large pitching moment and increase in forward velocity with increasing angle of attack for many powered configurations (Ref 1: pp13). Once special HSV considerations are accounted for in

the architecture, past experience and practice may be applied in a more conventional manner.¹⁰

There are a host of factors (called "governing properties" by STI) in hypersonic flight that may be considered key control issues.¹¹ Included among them is the precise longitudinal control required during rapid accelerations and decelerations when equilibrium conditions are difficult to identify. The first goal of proposed future research will be to investigate whether classical control strategies will result in architectures where gain scheduling is feasible.

For example, hypersonic climbout may include transition at Mach 3 to maximum dynamic pressure, then a Mach 6 climb at reduced dynamic pressure due to ramjet limits.¹² This may be loosely interpreted as flying through configurations P1BP, P2BP, P3BP, P4BP, P10BP, P4CP, P10CP, and P16CP of Reference 7 (See Appendix D). The control architectures to be developed will be tested on the above configurations for powered ascent, and those developed for descent will similarly be applied to successive points for unpowered descent. (If cooling totally dominates descent the flying qualities issues may be secondary to precise automatic control of flight path trajectory).

The second goal of proposed future research is to investigate the effects of control limiting and time delay on ascent and descent trajectories. This will be especially important for powered hypersonic flight where extreme sensitivities are anticipated between engine and airframe.¹³ The third goal will be to apply some of the new measures of MIMO stability to the above configurations.^{14,15} Finally, a pilot-in-the-loop analysis is proposed based on the Neal-Smith models as modified for super-augmented vehicles.¹⁶ One promising approach is the relationship between closed-loop extended bandwidth and pilot rating, which may be used to anticipate hypersonic flying qualities "cliffs" (Ref 16: pp 37).

This report reviews the HSV bare airframe dynamics for one generic, but not necessarily representative, HSV configuration. Two "minimal SCAS" strategies are then discussed for this configuration. Although the resulting architectures are multi-input, multi-output (MIMO), an attempt is made to close successive loops in a meaningful way so that insight that relates key variables is not lost. Classical analysis tools are used where appropriate, including "system

surveys" using SIGGY plots. Modern analysis tools are used to discuss MIMO uniform stability for hypersonic control laws, and the inclusion of time delays and control limiting is discussed. Finally, pilot-in-the-loop analysis is proposed for the "minimal SCAS" architectures and recommendations are made for future work.

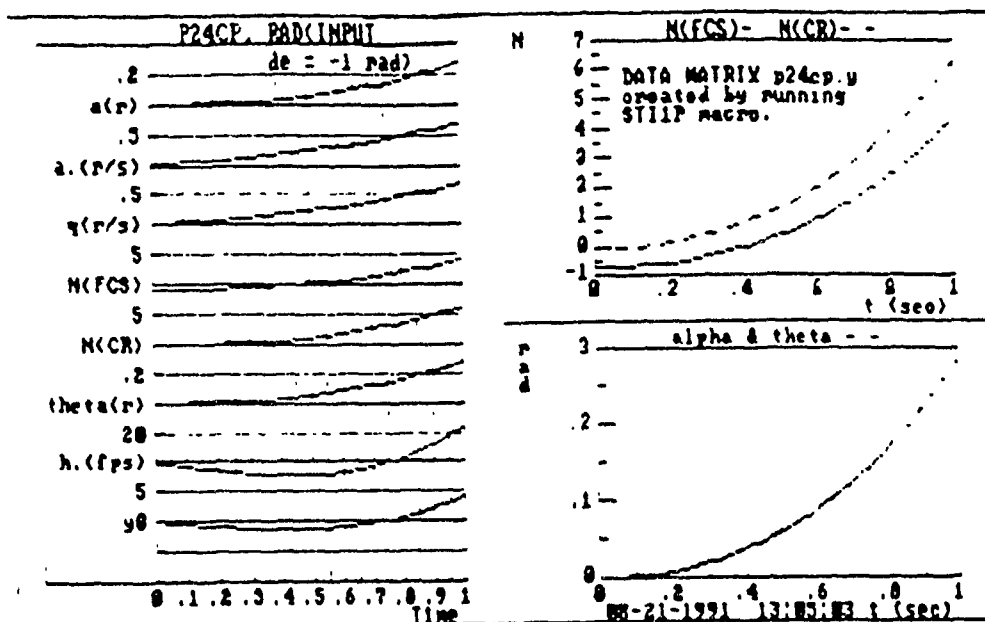
This introduction may best be closed by a quote justifying the concept of systems analysis: "Consideration of these idealized systems shows the ultimate performance approachable by some practical system or, by way of contrast, reveals directions in which it would be unprofitable to proceed." (Ref 10:pp 419)

2. AIRFRAME DYNAMICS

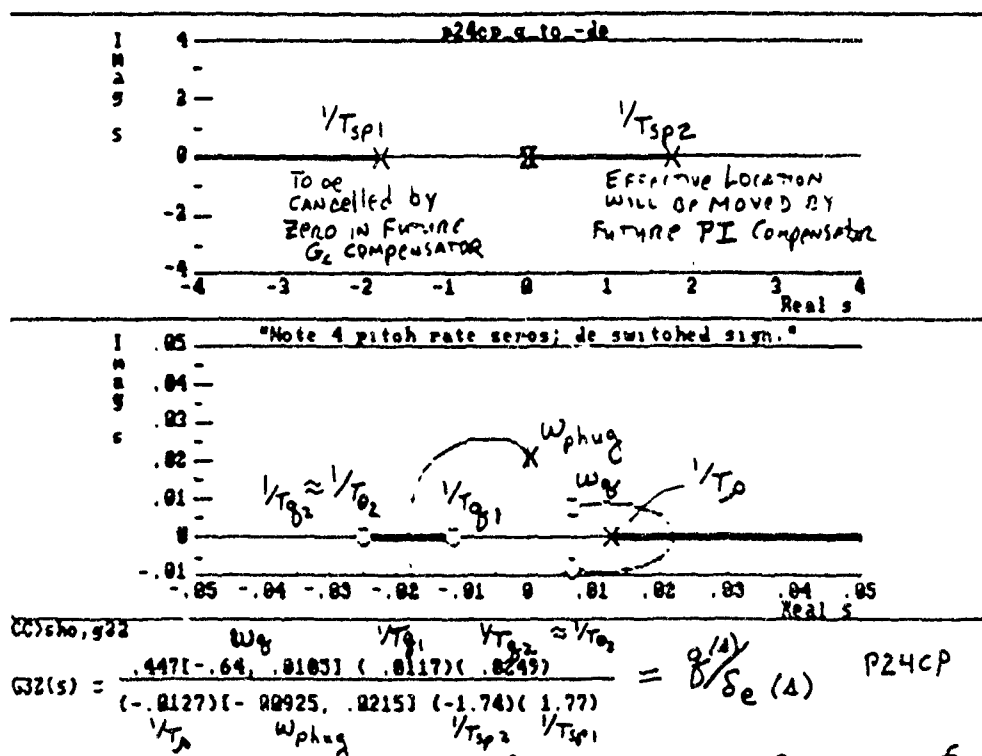
Although it is difficult to define equilibrium for a hypersonic vehicle, analysis of small perturbation models provided by Rockwell (Ref 7) show that open-loop instabilities are encountered throughout the flight envelope, that single-axis loop closures may often lead to closed-loop instabilities (due to right-half plane zeros), and that long period HSV dynamics have more modes than the simple phugoid mode of conventional aircraft and that these modes are considerably more complex.¹⁷

The potentially adverse impact of these properties on flying qualities can be addressed using the "minimal SCAS" concept, which proposes a control strategy on the bare airframe that permanently changes the vehicle as presented to the pilot. The importance of the airframe dynamics then is primarily as a baseline for the minimal SCAS design.

HSV Equilibrium. The locations in the hypersonic flight envelope where the base airframe dynamics are described by linearized models of a generic HSV are described in Appendix D. Equilibrium at each of 36 points POWER ON and 36 points POWER OFF is defined as a constant acceleration at a particular flight path angle. Note that the aircraft in equilibrium has positive, constant acceleration and that this constant value increases with dynamic pressure, regardless of whether power is on or off. This is due to the decrease in trim angle of attack (and associated drag) at constant flight path angle, even as dynamic pressure



STATEecho, FIGURE 1a: OPEN-LOOP SIMULATION OF P24CP (PLOT GENERATED BY @STI_RPT1) MACRO



(PLOT GENERATED BY @STI_RPT2 MACRO)

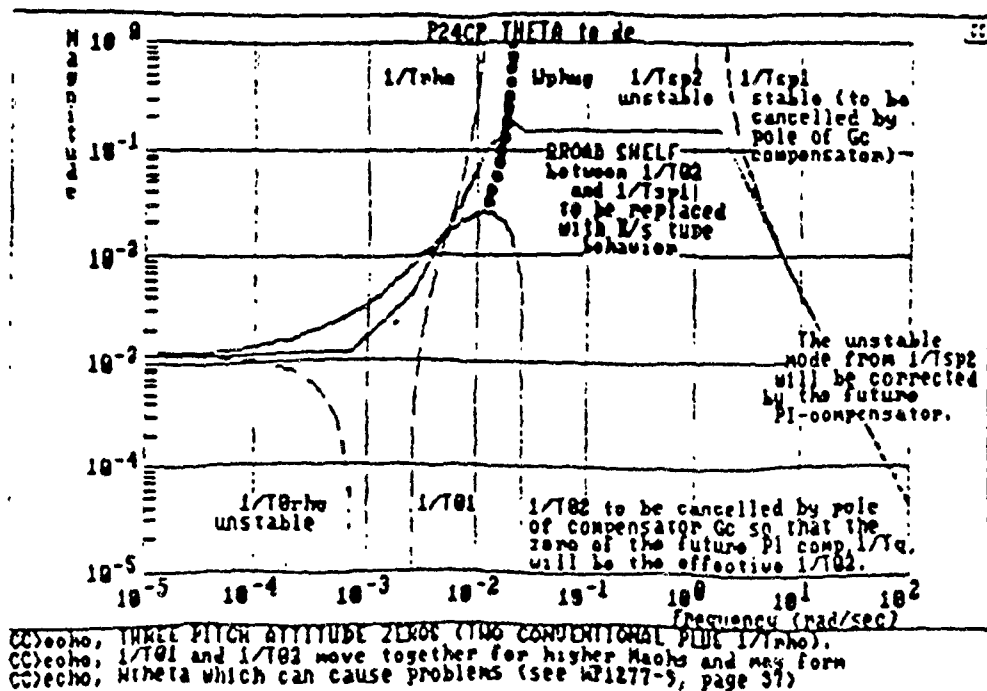
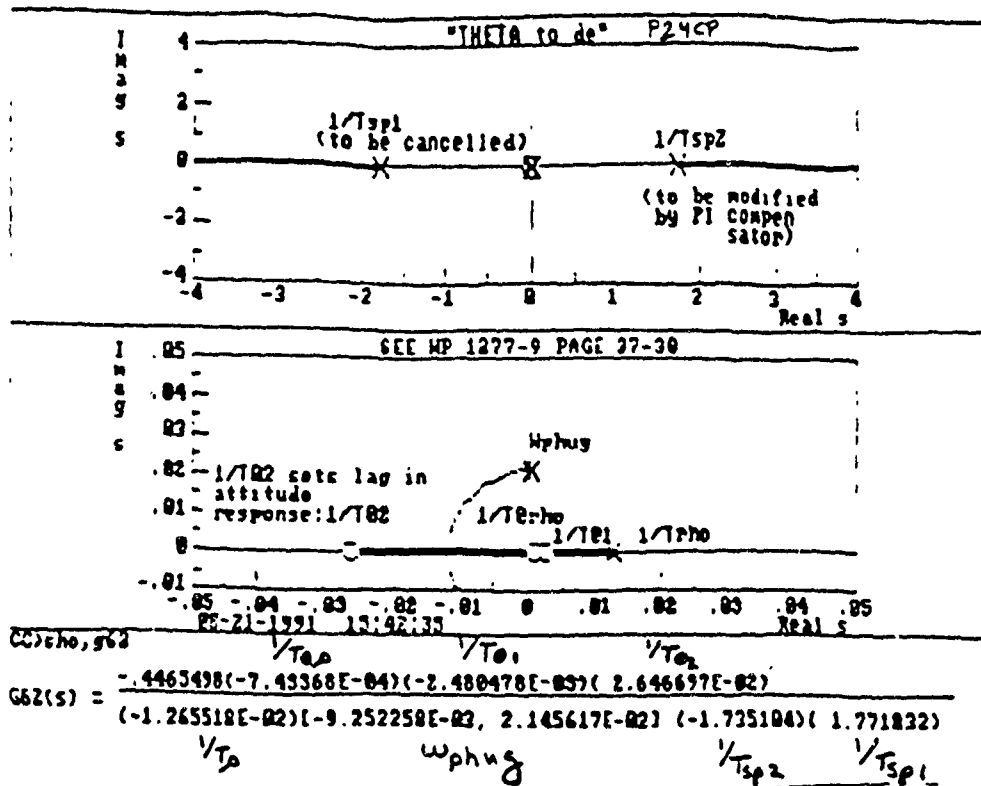


FIGURE 2. System Survey of Mach 20, $\bar{q} = 2000$ psf Powered HSV Configuration P24CP θ/δ_e Transfer Function

3. MINIMAL "SCAS" STRATEGIES

Because of the sensitivity of thrust/airframe coupling to inlet flow (i.e. to α , β , and Mach), uncommanded fluctuations of angle of attack and sideslip should be minimized. The initial loop closures that effectively determine the aircraft as perceived by the pilot, or minimal "SCAS", thus should minimize the effect of inlet flow variations. Past experience has shown that for robustness the system should be simple in concept and implementation so that it is potentially more effective in handling the inevitable uncertainties of HSV design.^{19,20} Finally, the evaluation of minimal "SCAS" strategies should be tested using flying qualities metrics that are applicable to high order control systems and to systems where the stability derivatives are not known precisely (the Neal-Smith and the bandwidth criteria, for example, satisfy these requirements).

The two minimal "SCAS" strategies examined in this report are the rate command attitude hold (RCAH) and the angle of attack command and hold (AAH) systems. If normal HSV operation consists of additional loop closures as presented to the pilot (path tracking command, for example), the minimal SCAS should still provide adequate flying qualities in the event of automatic system failure or override by the pilot. Minimal SCAS analysis will include, therefore, the closure of a path tracking loop in proposed future research.

3.1 SUPER-AUGMENTATION IN PITCH.

This augmentation scheme has been a prime topic of investigation by Systems Technology Inc (STI). The systems approach advocated by STI²¹ has matured and advanced to the point where task-tailored flight controls are feasible.^{22,23} The concept is currently being applied to selected HSV configurations of Appendix D (e.g. P24CP). The advantage of this approach over emerging multivariable control strategies^{26,27} lies in its simplicity, insight, and foundation based on past successes (Space shuttle, X-29).

The fundamental idea of superaugmented pitch control for an unstable airframe (i.e. one which has an unstable short period root ($1/T_{sp2} < 0$) is to replace the zero ($1/T_{\theta 2}$) controlling lag in the attitude response with a zero ($1/T_q$) of a PI-compensator (see Figure 3a). The goal is to change the broad, level "shelf" in the bode plot of q/q_c (known to cause flying qualities difficulties and illustrated in the SIGGY plot of Figure 2) with a frequency response having a more "flyable" K/s type of shape (as shown in the Bode plot of Figure 3).

The compensator G_c in Figure 3 helps to do this by cancelling the airframe $1/T_{\theta 2}$. It also cancels the airframe stable short period root $1/T_{sp1}$ so that the resulting root locus will have dominant second order roots (see the root locus in Figure 4) where dominant frequency and damping may now be chosen by setting K_q in the PI-compensator.

For HSV flight control this scheme is complicated by two difficulties. First, the lower frequency zeros of q/δ_e may be non minimum-phase (i.e. in the right half plane as in Figure 2). This will result in tracking performance as shown in Figure 4. This behavior is only safe if a pilot is continually attentive in the control loop. Other uncertainties in design, moreover, including the need for automatic control and the inherent lack of safety associated with unstable modes in flight, however slow they may be, justify the search for a stable minimal "SCAS".

The second difficulty is due to the rotation of the vertical, making the pitch reference θ different than the integral of pitch rate, q/s . Reasons for preferring θ to q/s are documented in Reference 25 (page 3). Figure 3b shows the change in the resulting transfer functions. Although the zeros of q/δ_e are no longer the zeros of θ/δ_e , standard superaugmentation approximations apply in the mid-frequency range (Ref 25: pp 3). The effect of non minimum phase zeros is now aggravated, but essentially of the same character as before in that they primarily affect long period instabilities. Appendix G discusses the rotation of the vertical and proposes some corrections to Reference 24 (STI Working paper-1277-8).

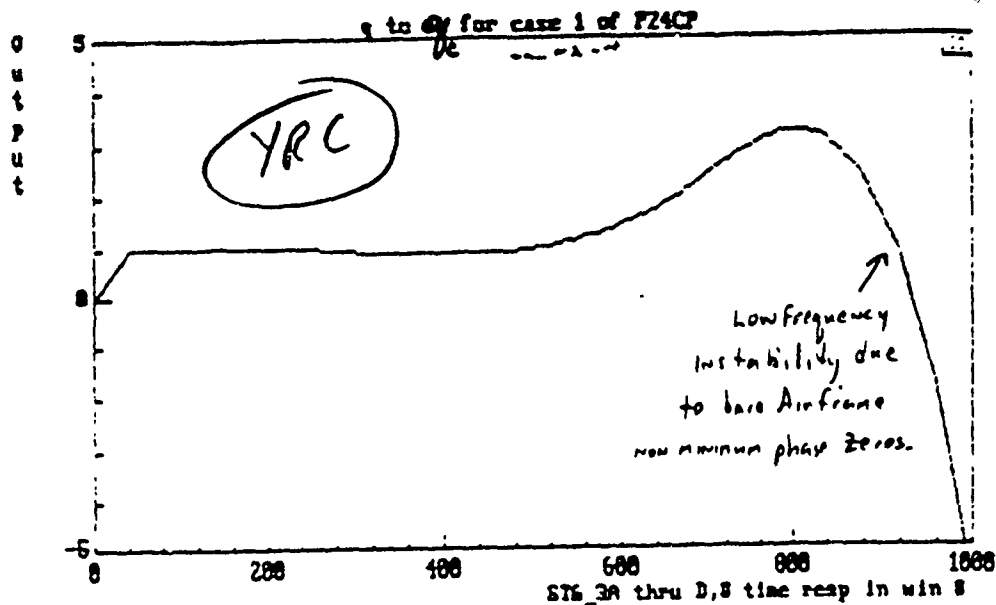


Figure 4a. TRACKING PERFORMANCE q/\dot{q} FOR SYSTEM OF Figure 3c.

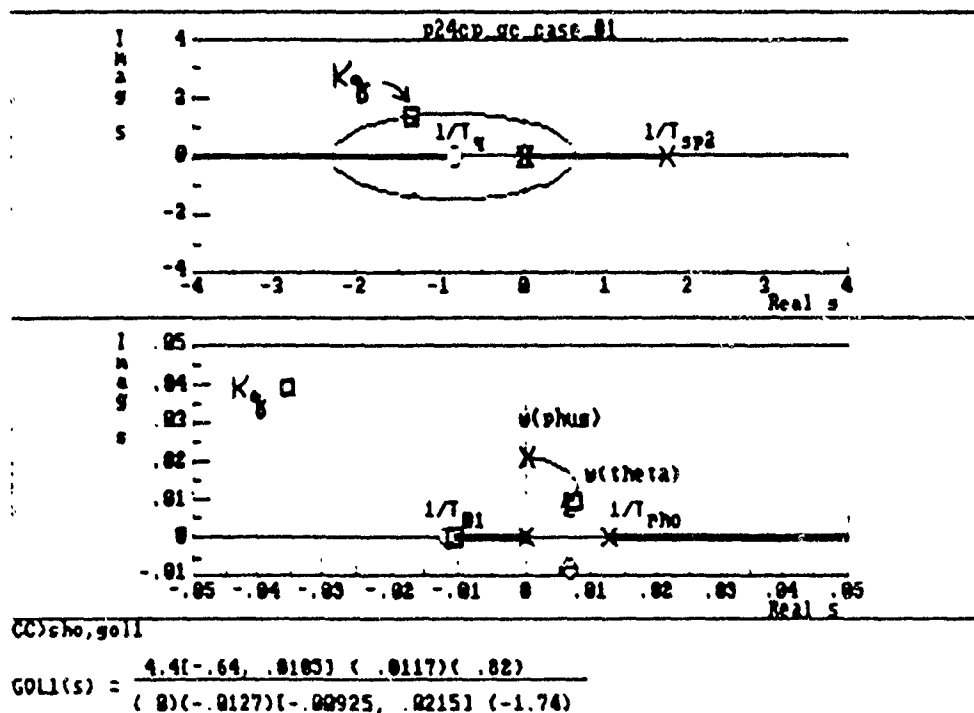


Figure 4b. ROOT LOCUS OF COMPENSATED PITCH LOOP SETTING DOMINANT POLE LOCATIONS

3.1.1 Closing the Velocity Loop.

One promising approach to remove the non minimum phase zeros of $q/\delta e$ is to close a velocity loop around the pitch loop. This is shown in Figure 5, which also includes a root locus survey for the $V/\delta e$ (velocity to thrust) transfer function for HSV configuration P24CP (Mach 20).

The goal is to select compensation G_v for the velocity loop so that the zeros of the pitch loop are minimum phase. Throttle to thrust dynamics, an important consideration, were not examined during this research period since we were not successful in determining a G_v compensator that resulted in minimum phase zeros for the pitch loop, a necessary first step.

The system of Figure 5 implemented a compensator of the form

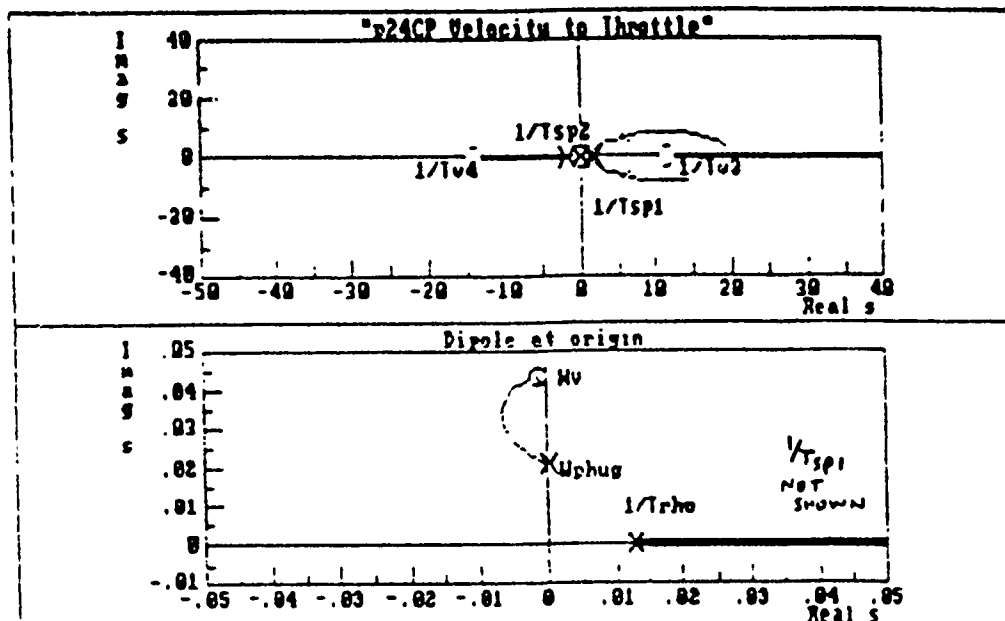
$$G_v = K_v (1 + T_v s) / (1 + T_t s)$$

Gain K_v and lead T_v were varied to study their effect in the pitch loop pole/zero locations. The plot at the bottom of Figure 5 shows the variation in the zeros of the pitch attitude loop as pure gain ($G_v = K_v$, $T_v = T_t = 0$) is varied from zero. It is interesting that $1/T_{\theta 1}$ and $1/T_{\theta 2}$ vary rapidly as K_v is varied in the velocity loop, coupling into a complex pair and then returning to the real axis. One of the zeros approaches a limit; the other travels to the left indefinitely.

The troublesome zeros, however, are the non-minimum phase oscillatory zeros, which become real as K_v is increased and approach limits of 0 and 0.3 respectively. Gain, then, is not enough to move these into the left half plane. No values could be found using a trial and error approach that resulted in minimum phase zeros. This is an area where further research is required.

Of course the poles of the pitch attitude loop also migrate as gain K_v is varied, but this is not shown in Figure 5. The stable short period pole moves indefinitely to the left on the real axis, the unstable short period pole couples with the density gradient pole becoming oscillatory, and the phugoid poles migrate vertically to a more oscillatory limit. The super-augmentation strategy, however, should still work on the resulting pitch loop.

In the next section an argument will be made for using angle of attack stabilization (AAH). It will be seen that, although AAH offers many advantages, it suffers from the same difficulty due to these non minimum phase zeros.



CCScho, g02

$$G83(s) = \frac{-3.88[.0261, .0431](-11.7)(14.2)}{(-.0127)[-0.00925, .0215](-1.74)(1.77)} = \frac{\Delta V(z)}{\delta_e(z)}$$

ω_v $1/T_{v3}$ $1/T_{v4}$ $1/T_{sp1}$ $1/T_{sp2}$

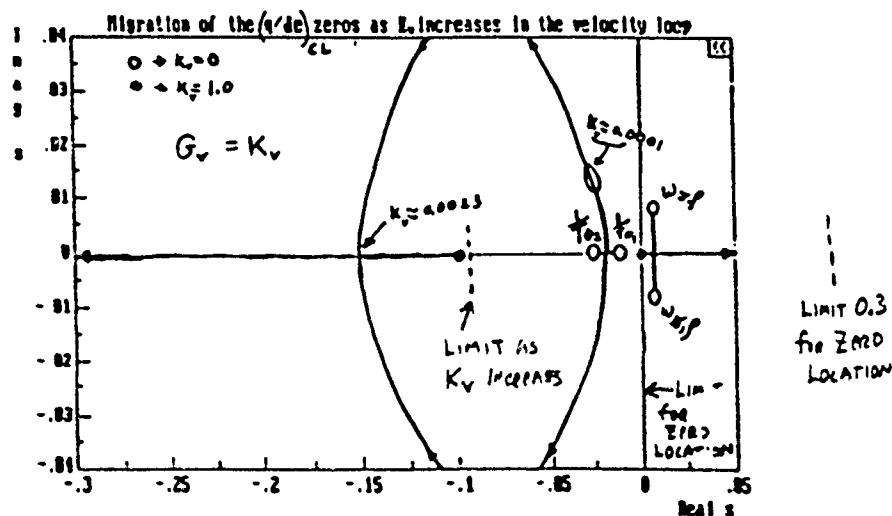
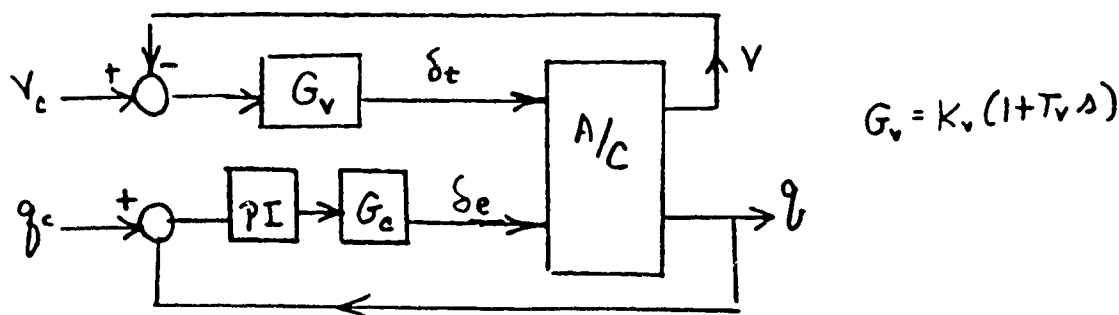


FIGURE 5. (a) VELOCITY TO THROTTLE ROOT LOCUS SURVEY FOR MICH 20 P24CP
 (b) MULTI-LOOP STRATEGY WITH VELOCITY LOOP CLOSED
 (c) PITCH LOOP ZERO MIGRATION AS K_v IS VARIED IN VELOCITY LOOP.

3.2 AUGMENTATION STRATEGIES USING ANGLE OF ATTACK

There are three reasons for considering angle of attack stabilization for hypersonic vehicles. The first is that alpha-stabilization presents a relatively conventional and stable vehicle to the pilot (Ref 10: pp 443-446). It further stabilizes the phugoid mode which, though usually unimportant in conventional aircraft, may be significant in HSV aircraft susceptible to phugoid coupling with the density mode (Ref 17). This stabilizing characteristic, however, is very similar to that provided by super-augmented pitch rate command and stabilization systems.

Since alpha-stabilization tends to maintain relative wind constant, fluctuations of HSV thrust should be minimized. Such fluctuations cause changing forces, moments, and possibly coupled airframe oscillations, all detrimental to controlled HSV flight. Moreover, it is possible that angle of attack must be precisely controlled due to envelope-limiting flight corridors. In this area alpha command and stabilization should control inlet flow conditions relative to the atmosphere better than a pitch rate command, attitude hold (RCAH) system. Moreover, there are rotation of the vertical effects causing horizon reference θ to be different than the integral of pitch rate, q/s , which a RCAH system must consider.

Alpha-stabilization will also reduce any roll-induced α to β coupling, critical for HSV flight to avoid beta-induced motions, heating, engine/airframe coupling, and adverse effects of engine unstart. A pitch rate command system may aggravate this type of sideslip coupling to angle of attack.

The downside of alpha-stabilization is the sensor suite required, which may be unavailable for practical HSV configurations, and the susceptibility of such systems to control limiting at high gain. The sensor difficulty may be alleviated by computing alpha and its derivative with a state observer or estimator. The pilot, who will probably not be flying relative to an actually observed horizon, would receive this information via an appropriate heads-up display (HUD). Command limiting may be required to avoid excessive external surface deflection rates.

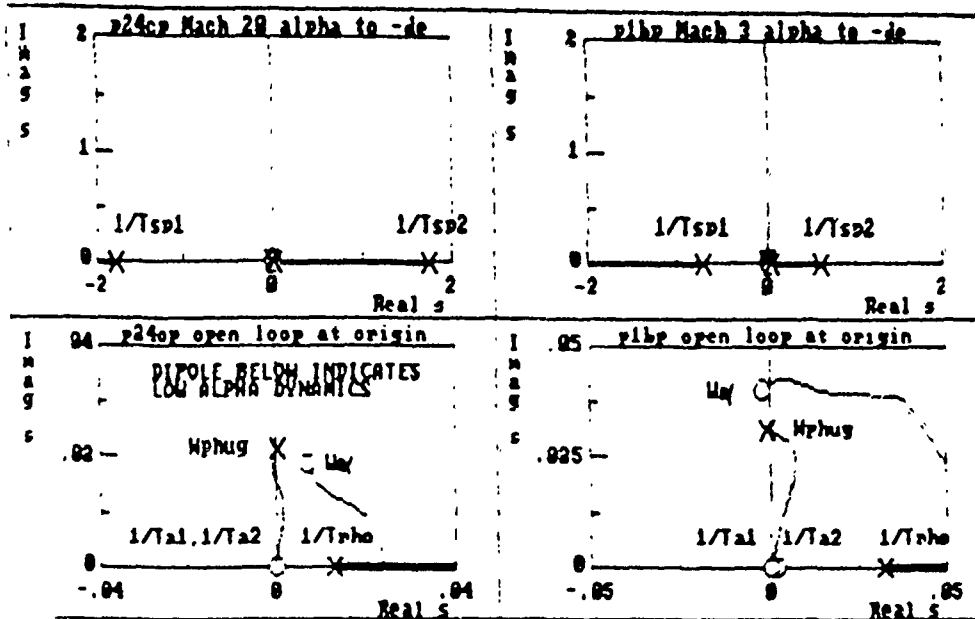
3.2.1 Airframe Angle of Attack Response.

Root locus surveys for two widely varying HSV configurations, P24CP (Mach 20, $q_{bar}=2000$ psf), and P1BP (Mach 3, q_{bar} 300 psf), are shown in Figure 6. The close proximity of the dipole ω_α and ω_{ph} for both configurations indicates little or no α -dynamics at low frequencies.

Angle of attack super-augmentation for an unstable aircraft will proceed as discussed in Section 3.1 for pitch rate super-augmentation. A PI-compensator for the Mach 20 configuration P24CP is chosen to provide reasonable bandwidth, and a Gc compensator is chosen to provide a broad frequency of K/s type of response, as shown in Figure 7.

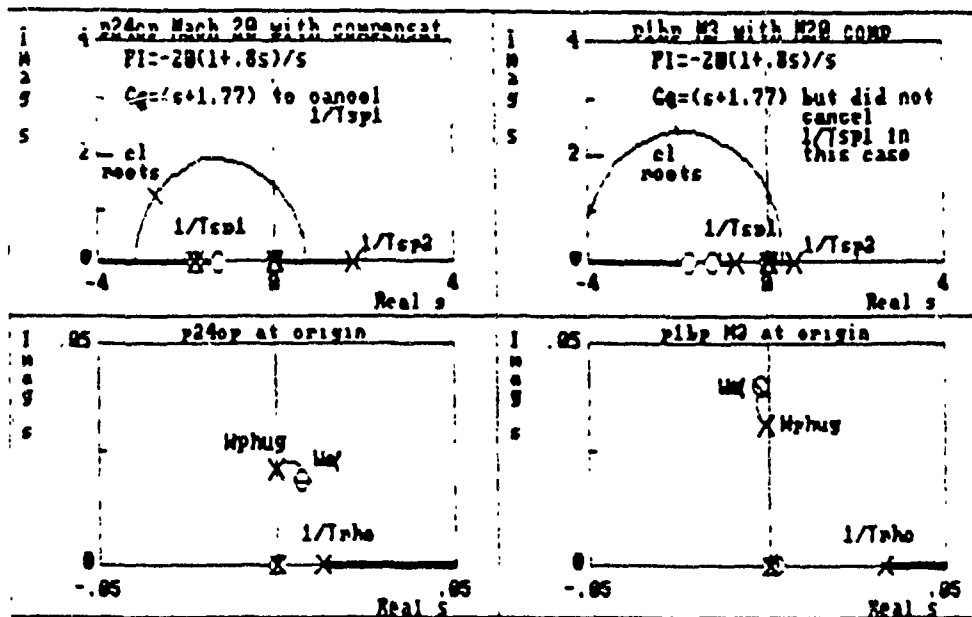
Figure 8 shows the tracking response of both systems using the compensator designed for the Mach 20 configuration. Robust behavior is apparent for the initial five seconds of response (what the pilot would "see"), and the long period instabilities appear many minutes after the step command.

Angle of attack stabilization thus illustrates the same non-minimum phase behavior as pitch rate stabilization. It is expected that a second loop closure will remedy this problem, but a successful solution was not found during the summer research period. It is proposed that this multiple-loop stabilization be accomplished in follow-on research.



$G_{\alpha} = \frac{.0027(0)(-8.1E-5)(-.347, .02)(52.3)}{(-.0127)(-.00925, .0215)(-1.74)(1.77)}$
M20 $\frac{1/T_{sp1}}{\omega_{PH}} \frac{1/T_{sp2}}{1/T_{sp1}}$

$G_{\alpha} = \frac{.00617(0)(-180E-5)(.059, .04)(352)}{(-.0324)(.021, .0314)(-.6)(.72)}$
M3 $\frac{1/T_{sp1}}{\omega_{PH}} \frac{1/T_{sp2}}{1/T_{sp1}}$



M20

M3

Figure 6. Angle of AHACK STABILIZATION AT MACH 20 AND MACH 3
 USING THE MACH 20 COMPENSATOR
 $PI = -20(1+.8s)/s$ $G_c = (s+1.77)$

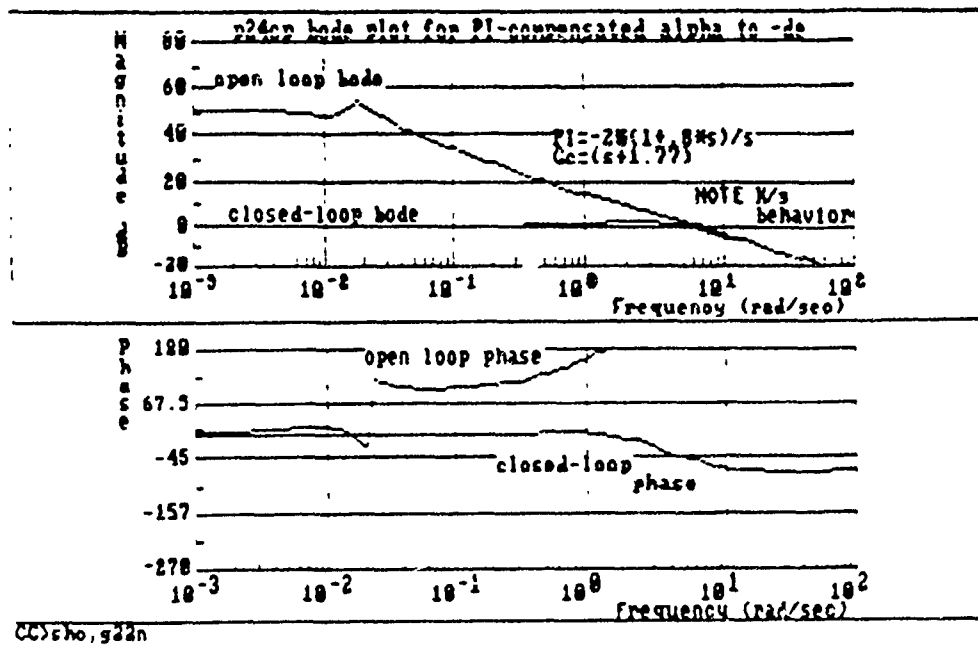


Figure 7. Open Loop Bode Plot for Mach 20 P24cp with Compensation in Alpha Forward Loop $PI = -20(1 + 0.8s)/s$
 $G_c = (s + 1.77)$

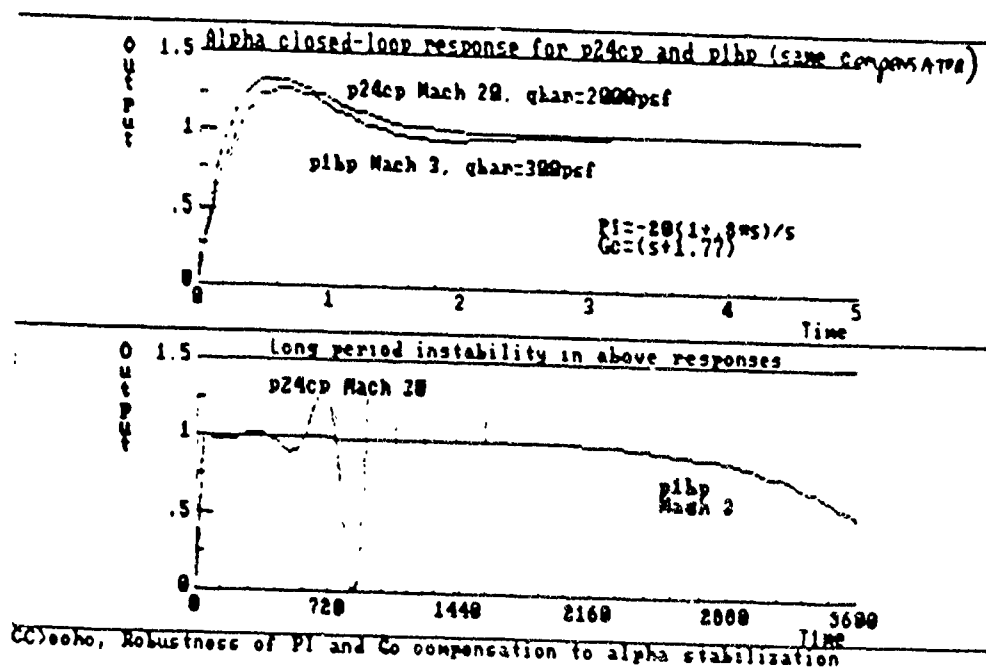


Figure 8. Closed-Loop Time Responses illustrating Short Term Robustness and Long Term Instability of Mach 20 and Mach 3 HSV Configurations using Mach 20 Compensator

4. SPECIAL CONSIDERATIONS

4.1 CONTROL LIMITING.

Since an augmented, unstable aircraft tends to return to its unstable modal behavior whenever control limiting occurs (see Ref. 9), it is essential that any control strategy be tested by monitoring control deflection and rate as a function of open loop gain (i.e. system bandwidth). Rate limited actuators may be simulated by first order lags using transfer functions or as augmented state equations in state-space representation. Proposed research into this area will use both schemes. CC Macro @ST15 has been developed as a tool to investigate control limiting. CC Macro ST16 analyzes the resulting system made by @ST15.

4.2 TIME DELAY

The ill effects of time delay on flying qualities are well known. Although such "lumping" of time delay into one point of the control system is not rigorous, it does effectively convey the loss of performance and tendency towards instability. Proposed research will vary time delay between 0.0 and 0.2 seconds and investigate its effects on loop performance.

4.3. MULTI-LOOP SYSTEM STABILITY

Figures of merit for the stability of multi-variable flight control systems are not yet well established in industry. However, much theory has been developed in this area that is promising and which should be investigated. A distinction may be made between "nominal" and "robust" stability. Current flying qualities guidelines, as given in MIL-STD-1797, emphasize "equivalent" systems to determine nominal stability measures of damping and time constants. These systems may be poor indicators of stability²⁹ and be totally inappropriate when unstable modes are approximately cancelled by non minimum phase zeros.

Multi-loop stability margins and open-loop bandwidth are used to specify MIMO robust stability margins. ^{14,15} It is not known how these new measures will be used to affect HSV design. These topics will be covered in proposed future research, and the control strategies

developed will be checked for both nominal and robust stability using both independent and uniform measures.

4.4. FLIGHT PATH ANGLE RESPONSE

It is assumed that flight path control will be a dominant consideration in HSV flight. A path loop thus is anticipated to be closed around the minimal SCAS regardless of the type of control strategy chosen. The path bandwidth is an important determination of the type of tasks which may be safely accomplished.²² For HSV flight it will be important to know whether configurations are phase margin limited (typical) or gain margin limited (oscillation prone). Guidance for applying bandwidth criterion is explained in MIL-STD-1797 (see also page 31 of Reference 16). Bandwidth metrics will be proposed for future research.

4.5. PILOT-IN-THE-LOOP ANALYSIS

The revised boundaries of the Neal-Smith criterion (Ref 16: pp36) predict pilot-in-the-loop performance ratings given the aircraft and control system transfer functions. Unfortunately, these boundaries are not known for any task involving HSV flight. It is clear, however, that aircraft sensitivity to task demands or pilot technique is an important factor determining flying qualities "cliffs". This sensitivity may be approximated by plotting predicted pilot rating over a range of realistic task bandwidths (Ref 16: pp 37). Rapid deterioration in rating is indicative of excessive sensitivity and latent poor handling qualities. It is proposed that HSV sensitivity to task bandwidth be determined in future research.

5. CONCLUSIONS AND RECOMMENDATIONS

The major conclusions and recommendations of this report are listed below:

(1) The super-augmented minimal SCAS strategy is viable for controlling either pitch rate or angle of attack and presents a "flyable" aircraft to the pilot in the short term. It is recommended that these strategies be evaluated for multi-loop nominal and robust stability, control limiting, the inclusion of time delay, and pilot-in-the loop behavior.

(2) Angle of attack super-augmentation may be preferable to pitch rate control due to its direct control of inlet flow to minimize thrust/airframe coupling. It also reduces α to β coupling while rolling and the chance of unstart. It is recommended that the difficulties of implementing this scheme using achievable sensor dynamics and computation be investigated.

(3) The primary difference in the longitudinal modes between HSV dynamics and conventional unstable aircraft configurations lies in the low frequency instabilities caused by the height, density gradient, and phugoid coupling. The rotation of the vertical and centrifugal relief are very long period phenomena that should not affect control law operation in the short term. It is recommended, however, that the long period instabilities (non-minimum phase zeros) be eliminated by an appropriately compensated velocity loop.

(4) Controls analysis packages were found to be indispensable in working with the large number of HSV generic configurations. It is recommended that macros (and/or M-files) be developed to provide rapid review and access to design data and strategies.

FOR REFERENCE PAGES
SEE REPORT BY 'Dr. Daniel J. Biezd
Vol.#5

A STUDY ON THE EFFECTS OF TEMPERATURE AND
GAS COMPOSITION ON THE MICROSTRUCTURE/MORPHOLOGY
OF CVD SiC IN THE MTS/H₂ SYSTEM

Peter Brown
Graduate Research Associate

Abstract

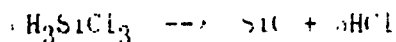
Silicon carbide was deposited onto graphite substrates in a vertically-oriented hot-walled reactor using a processing technique known as chemical vapor deposition (CVD). The precursor gas used in this study was methyltrichlorosilane (CH_3SiCl_3 = MTS). Hydrogen was acted as the carrier gas. Experiments were conducted to study the effects of temperature and inlet composition on the morphology of the deposited coating. It was shown that the microstructure and morphology of the coating do depend heavily on the temperature and the composition of the inlet gas stream. Further experimentation should be done to understand these dependencies more thoroughly, as well as the silicon carbide's dependence on the total system pressure and the mass flow rate into the reactor.

Introduction

A wide variety of gaseous precursors, temperatures, pressures and inlet compositions (as well as reactor geometries and configurations) have been employed in the past to deposit silicon carbide using chemical vapor deposition (CVD). Silicon carbide remains an important material with many potential applications as a structural material due to high oxidation and wear resistance. Another possible application is as an intermediate material between the fibers and the matrix of a composite material. For instance, in the case of high thermal mismatch between the fiber and the matrix, a third material may be deposited onto the fibers which has intermediate properties between the fiber and matrix. Silicon carbide is inherently a relatively hard and strong material. The mechanical properties, however, are dependent on the microstructure and morphology of the deposit. The microstructure and morphology are in turn dependent on the deposition parameters chosen to perform the CVD process (for example, temperature, pressure, inlet gas composition and gas flow rate). Knowledge of these dependencies would greatly minimize trial and error needed to determine the parameters for the successful deposition of coatings with the desired mechanical properties.

Methyltrichlorosilane (CH_3SiCl_3 , MTS) offers perhaps the easiest route to depositing a stoichiometric $\beta\text{-SiC}$, since within each molecule of MTS there is one atom each of carbon and silicon. In this particular study, MTS was used as the precursor gas and hydrogen was utilized as the carrier (diluent) gas. The basic reaction taking place here is

assumed to be the following:



where the hydrogen gas acts only as a carrier gas.

When MTS is heated to high enough temperatures, the vapor decomposes and deposits SiC on all surfaces inside the furnace which are above a critical temperature. Hence in a hot-walled reactor, whereby the substrate is heated by the walls of the furnace, SiC also deposits onto the furnace walls. However, in a cold-walled reactor where the substrate is typically heated through resistance heating or rf induction heating, the walls of the furnace generally do not reach high enough temperatures for deposition to take place on them.

The objective of this study was to examine the effects of various deposition parameters (i.e. temperature and inlet gas composition) on the overall deposition rate and surface morphology of the silicon carbide grown by CVD in the MTS/H₂ system. Also, another desirable outcome of this research project was to identify certain regions along the length of each substrate in which a particular morphology dominated the structure of the coating.

Experimental Procedure

Figures 1 and 2 show a schematic of the CVD system¹. There are five basic steps involved in performing the deposition run: (1) physically assembling the reactor with all the plumbing and pressure gauge connections, and installing the substrate (in our case the substrate is a 3 mm diameter graphite rod, eighteen inches in length, supplied by the

Grabner Company in Dayton. (1) setting the deposition parameters: (2) coating the substrate: (3) shutting down the system: and (4) removing the substrate from the furnace and disassembling the system. A full description of these steps (how to assemble the furnace, look up the necessary plumbing, gauges, etc) is however not within the scope of this study. Therefore only a discussion of the deposition parameters studied will be presented.

Three sets of deposition parameters were studied. The first set is used primarily in this study as a reference point; the other two conditions can be used to infer the relationships between the temperature and inlet gas composition on the deposition rate and morphology of SiC. The following table gives the experimental conditions for each trial.

<u>Trial</u>	<u>Temp(C)</u>	<u>Pressure(torr)</u>	<u>H₂/MTS</u>	<u>TFR(sccm)</u>
A	1300	100	10:1	330
B	1200	100	10:1	330
C	1300	100	2:1	450

Each trial (A, B, C) was conducted twice in order to show some sort of experimental repeatability. Each experiment was conducted for thirty minutes. The H₂/MTS column represents the molar (and volumetric) ratio between hydrogen and the MTS vapor at the inlet of the furnace. "TFR" is an acronym for total flow rate, as measured in standard cubic centimeters per minute. The amount of hydrogen gas flowing into the furnace

was held constant throughout all the trials at 300 sccm. Depending on the desired H_2 /MTS ratio, the MTS volumetric flow rate was altered accordingly. The temperature listed represents the set point of the furnace. The actual furnace used was a vertically-oriented tube furnace, eighteen inches in length. The temperature profile (experimentally measured¹) schematically looks like the following:

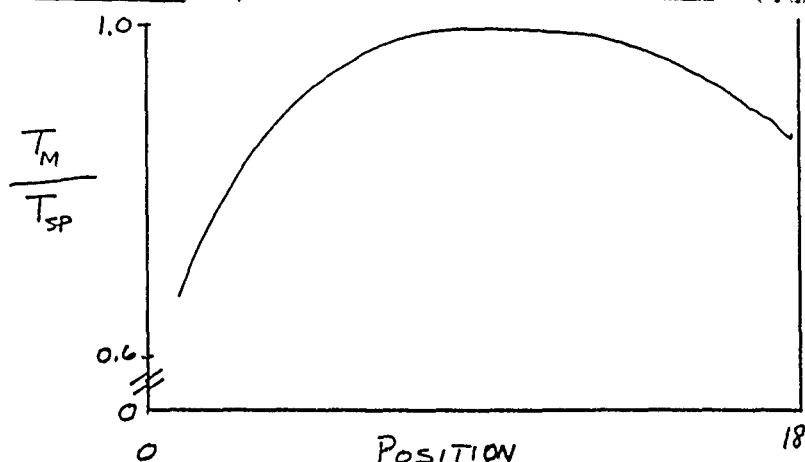


Figure 3.
Temperature profile schematic.

The vertical axis T_m/T_{sp} represents the measured temperature as a fraction of the set point temperature. Hence a value of 1.0 along the y-axis would indicate that the measured temperature was precisely the value of the desired set point temperature.

The horizontal axis on the figure represents the position along the longitudinal axis of the furnace at which the temperature was measured. A position of "0" represents the inlet of the furnace, and a position of "18" represents the position of the furnace 18 inches from the inlet, which happens to be the outlet in this case. Hence a position of "6" for example would represent a position six inches from the inlet of the furnace.

It would be appropriate to comment on this temperature profile.

it is not quite symmetrical in shape. There is an approximately 8" long temperature ramp neighboring the inlet, followed by a 26" isothermal section (roughly speaking), which in turn is followed by a 20" section over which the temperature is falling (albeit at a slower rate than at the inlet).

The substrate as mentioned earlier was a 3 mm diameter graphite rod, 18" in length. The substrate was hung from the top by some carbon yarn, in such a manner that the bottom end of the graphite rod was positioned precisely at the "0" position of the furnace (at the inlet).

At the conclusion of each experiment, the eighteen inch graphite rod was fractured into 0.75" lengths. Each fracture surface was in turn examined under the SEM to determine the total deposition thickness and the surface morphology of the SiC coating.

Experimental Results & Discussion

The next three figures (4, 5 and 6) depict the average deposition rate for trials A and B with respect to the location in the furnace. Figures 4 and 5 shows the data for both experiments under trial A. The peak deposition rate for trial A is around 1.3 microns per minute, and is located about 5" from the inlet. It is presumed that the reaction kinetics are the rate controlling step for the SiC deposition reaction from this data. The temperature at the inlet does not provide enough thermal energy to overcome the reaction barrier until the gases reach the 3-4" position. Throughout the hottest region of the furnace, there is a general decrease in the deposition reaction. This is probably due

to the depletion of the reactant gas, MTS. Even though the reaction is most favored at the higher temperatures, the consumption of MTS at points upstream of the "hot zone" (the 6" or so which is fairly isothermal, and also the hottest part of the furnace) does not allow a higher deposition rate within the hot zone.

Figure 6 shows a similar plot for the experimental conditions listed under trial B. The only difference between trial A and B is the set point of the furnace (1300 and 1200 °C, respectively). Two major differences between the two trials can be seen from these data. At a lower temperature, the overall deposition process is significantly retarded (the average rate is cut approximately in half by reducing the temperature 100 K). Also, the deposition profile for the 1200 °C case is shifted by about two inches downstream from the 1300 °C profile. That is, the initial deposition for the 1200 °C case did not start until a point about 2" further downstream than the 1300 °C case. This shift in location can be attributed to the shift in the point of the furnace which has a high enough temperature to start the reaction. Also, since the initial deposition for the 1200 °C case was lower, there was still some MTS left over to react further down into the hot zone, when compared to the 1300 °C case. Hence there are sections of the graphite rods which have more coating on them for the 1200 °C case than do the 1300 °C case sections from the same positions in the furnace.

No deposition profile was plotted for the trial labeled as C. The deposited SiC had such a non-planar surface morphology in certain sections that any attempt to measure the deposition thickness proved unsuc-

cessful.

The following four figures (7, 8, 9 and 10) show some of the SEM micrographs taken of the fracture surfaces of the three trials. Only four representative sections are shown here, although data are collected every 0.75" along the length of each graphite rod coated with SiC. The first figure (7) shows all three trials' fracture surfaces at the 3.75" position of the furnace. The A and C trials are very similar (same temperature, different inlet compositions), whereas trials A and B are significantly different at this point (different temperatures, same inlet gas composition). Trials A and C are fairly nodular at this point, and trial B appears to be much more whisker-like in morphology.

Figure 8 shows the SEM micrographs of the three samples at the 7.5" position. Here all morphologies are about the same roughness. It is interesting to note here that the trial C coating is not as thick as the trial B coating, although the trial C conditions have a much richer inlet concentration of MTS than the trial B.

Figure 9 show the same samples' fracture surfaces at the 28.25" position in the furnace. The most drastic change occurred in the trial C sample, where in less than 1" along the axis of the furnace the morphology transformed from fairly nodular to something completely composed of whiskers about 100-200 microns long and about 4 microns wide. Trials A and B underwent almost no morphological change within the same change in position in the furnace. It should be mentioned that according to the measured temperature profile, both the figure 8 and 9 micrographs originate from the isothermal section of the furnace. Hence

it appears the major difference in the conditions for figures 8 and 9 could be the composition of the gas next to the substrate. Not only does the inlet gas composition play an important role in the morphology of the SiC coating, but the gas composition neighboring the substrate also plays a critical role.

Figure 10 shows the fracture surfaces of the trials, located at the 11.25" position in the furnace. The morphologies of all three trials did not drastically change in this location compared to the data in figure 9. The deposition occurring under trial A conditions appears to be only $\frac{1}{4}$ that compared to the trial B conditions. Trial C remains to be composed of whiskers, albeit shorter whiskers than at the 8.25" position.

Figure 11 shows a rough schematic of the various morphologies found for each set of experimental conditions, as a function of position along the axis of the furnace. The drawings are not to scale, and were designed solely for a qualitative analysis of the morphologies. Also, it should be noted that the boundaries as drawn do not really appear as sharp distinct boundaries. Instead, the boundaries between two neighboring regions are usually spread over a 1" (or perhaps a little less) length of the substrate surrounding the drawn boundary. By comparing figure 11a and 11b, it can be seen that the central whisker region found in 11a has disappeared when the temperature of the furnace was decreased. Also, it can be seen that the deposition process has moved downstream in the lower temperature case when compared to the 1300 C case.

By comparing figures 11a and 11c, one can see by increasing the inlet concentration of MFS, the initial growth of whiskers has been suspended. In addition, for the case of the richer inlet MFS concentration, a new region appears, namely this 'shiny' section, which with the naked eye appears to be almost mirror-like in finish. Under the SEM this section also appeared much smoother than other sections of all trials. Another observation from comparing 11c to the others would be that the onset of whisker growth towards the outlet end of the furnace is initiated at a much earlier position for the trial C ($\approx 8"$) than for the trial A and B ($\approx 13"$). This is probably caused by the changing composition of the gas phase directly above the substrate.

Best Available Copy

CONCLUSIONS

Experiments were conducted in the MTS/H₂ system to study the effects of temperature and inlet gas composition on the morphology and overall deposition rate of SiC on a graphite rod. It can be generally noted that the temperature and the inlet composition both play important roles in the morphology and deposition rates along the rods' axes. At richer MTS concentrations, the morphology became almost planar in a small region towards the inlet, and changed drastically to a morphology composed almost completely of whiskers. At the lower temperature tested, the morphology did not change so dramatically compared to the higher temperatures, other than perhaps appearing slightly more planar. Additional data is required in order to make more precise statements about the effects of these parameters on the deposited thickness and morphology of the coating.

Best Available Copy

References .

1. Unpublished work of Jason Guth, WRDC/MILLN.

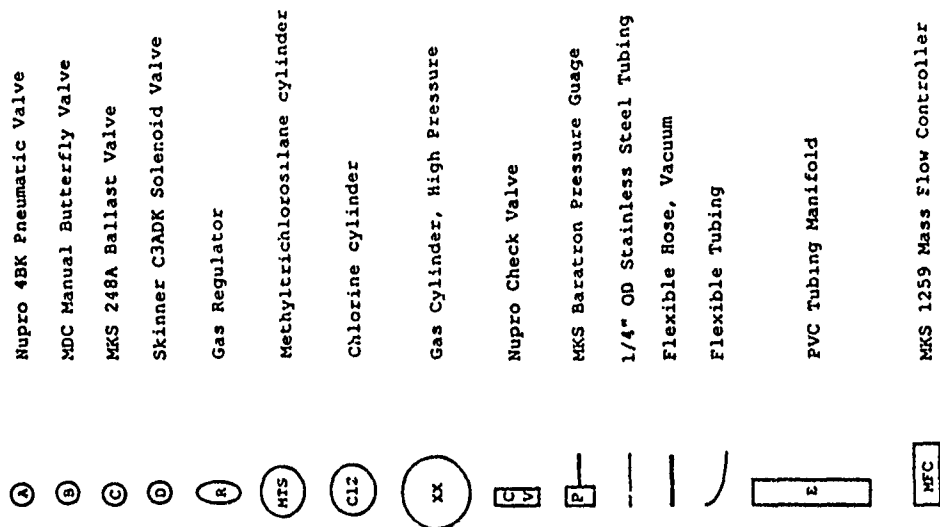
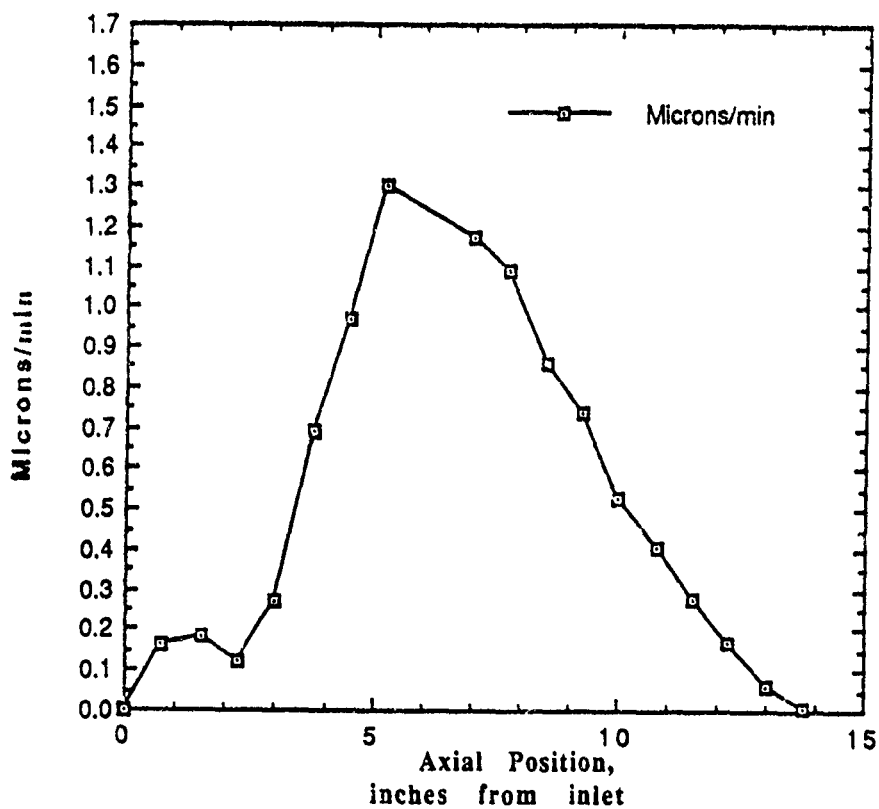
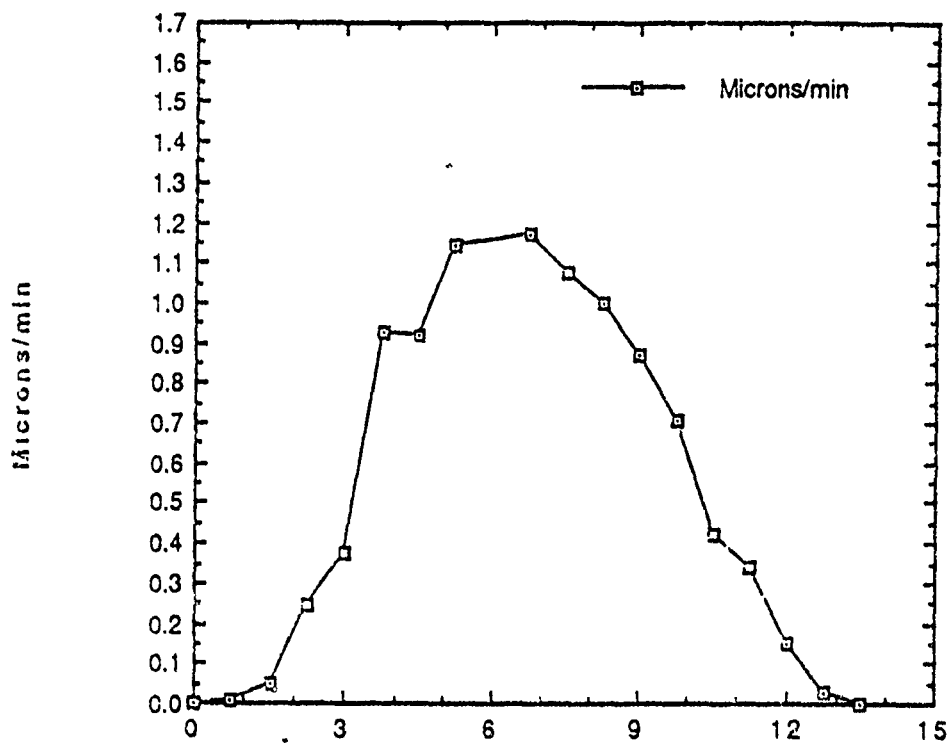


Figure 2. Gas flow schematic continued.

T=1300 °C, P=100 torr, H₂/MTS=10:1,
TFR=330 sccm



Figures 4&5. Deposition profiles for trial A conditions.

T=1200 °C, P=100 torr, H₂/MTS=10:1,
TFR=330 sccm

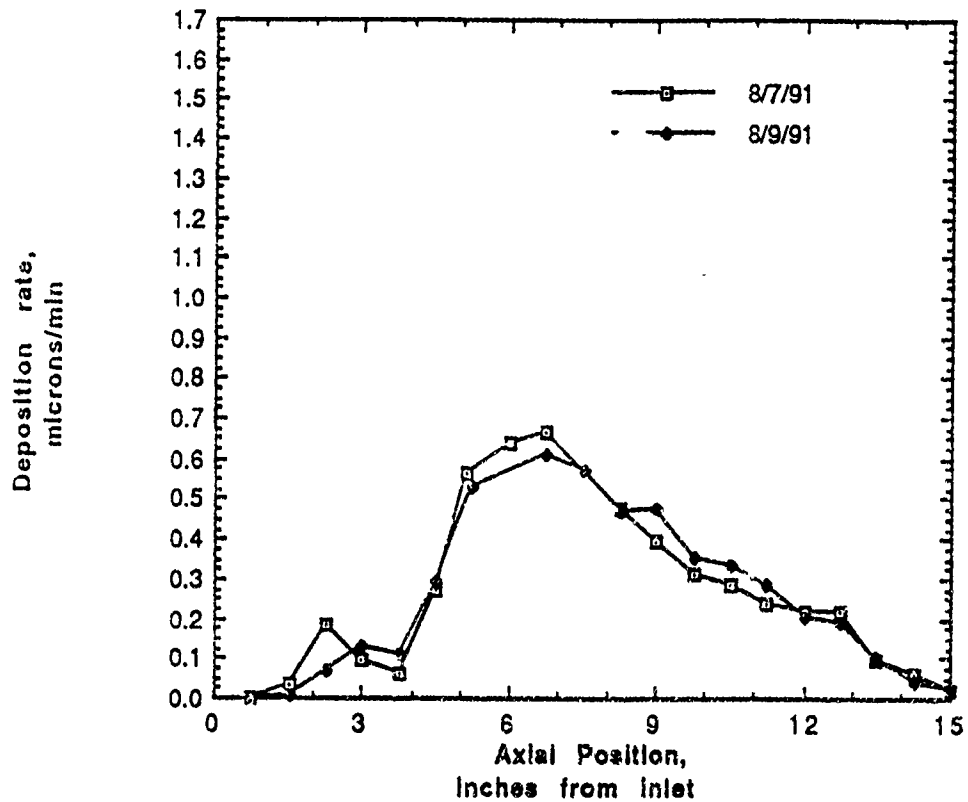


Figure 6. Deposition profiles for trial B conditions.

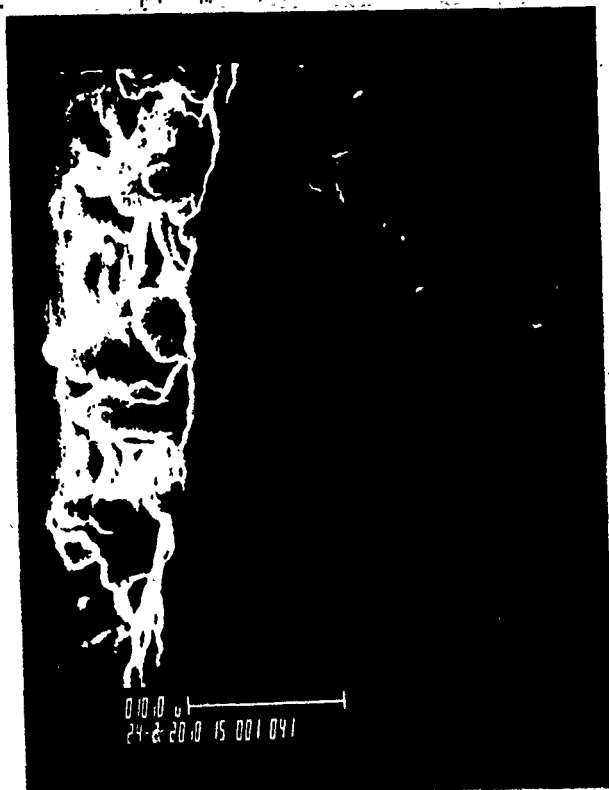


Figure 7. SEM micrographs for all three trials, $\theta \approx 3.75^\circ$ position.

Upper left:	trial B
Upper right:	trial A
Bottom:	trial C





Figure 8. SEM micrographs for all three trials, @ $\approx 7.50^\circ$ position.

Upper left:	trial B
Upper right:	trial A
Bottom:	trial C



Figure 9. SEM micrographs for all three trials, $\theta = 8.25^\circ$ position.

Upper left:	trial B
Upper right:	trial A
Bottom:	trial C



Figure 10. SEM micrographs for all three trials, @ 11.50" position.

Upper left:

Upper right:

Bottom:

Trial B:

Trial A:

Trial C:



$T = 1300\text{ }^{\circ}\text{C}$ $H_2/\text{MTS} = 10:1$

ϕ	SMALL WHISKERS	NODULES	WHISKERS	ROUGHER NODULES	VERY SMALL WHISKERS
11a.	1"	2"	4½"	6"	13"

$T = 1200\text{ }^{\circ}\text{C}$ $H_2/\text{MTS} = 10:1$

ϕ	SMALL WHISKERS	NODULES	WHISKERS
11b.	3"	4"	13½"

$T = 1300\text{ }^{\circ}\text{C}$ $H_2/\text{MTS} = 2:1$

ϕ	NODULES	SHINY SECTION	NODULES	VERY LONG WHISKERS	SMALL WHISKERS
11c.	1½"	3"	4"	7½"	12"

Figure 11.

Schematic of the "morphology zones" present for each trial.

THERMAL DEGRADATION OF PERFLUOROPOLYALKYLETHER FLUIDS IN THE
PRESENCE OF VARIOUS METALS

DAVID DRAKE

ABSTRACT: The thermal stability of Perfluoropolyalkylether (PFPAE) fluids (Krytox, Aflunox, Demnum, Perfluoropolydioxolane and perfluoropolyethylene oxides) has been investigated at different temperatures in the presence of three metals, 1018 steel, M-50 steel and M-50NIL steel. The preliminary results found 1018 steel to be much more reactive towards the PFPAE fluids in comparison to both M-50 and M-50NIL steel. It was also determined that, out of the fluids looked at Perfluoropolydioxolane was the most reactive in the presence of 1018 steel. The reactivity of the fluids was found to be as follows:
Perfluoropolydioxolane > Krytox > Perfluoropolyethylene oxides > Demnum > Aflunox.

INTRODUCTION: Perfluoropolyalkylether (PFPAE) fluids hold an interest as potential high temperature liquid lubricants as well as hydraulic fluids and base fluids for high temperature greases. The most important properties that perfluoropolyalkylethers exhibit are their thermal stability, oxidative stability and chemical inertness. Out of these the ability to resist oxidation at elevated temperatures is the most impressive. This outstanding property is attributed to the significantly higher resistance of the carbon-fluorine bond to oxidation compared to the carbon-hydrogen bond, which is, in turn, attributed to the electron deficient nature of the bond due to the strongly electron withdrawing fluorine atom (1). As a result there is a lack of electrons at the carbon-fluorine bond which provides for a poor site of attack by oxygen compared to the electron rich carbon-oxygen bond.

The performance of perfluoropolyalkylethers is dependant upon their molecular architecture. Whether a system is branched or linear and even its carbon to oxygen ratio has an effect on the properties mentioned above. As a reference the name as well as the structure of the five fluids looked at is on the following page.

EXPERIMENTAL: Three metals (1018, M-50 and M-50NIL) along with five PFP AE fluids were used in the following procedure. The metals, in the shape of flat disks and referred to as coupons were polished to a mirrored finish using sandpaper of grit 320, 400, and finally polishing paper. The coupons once weighed (to the nearest .01 milligram), were then placed in test tubes and to this PFP AE fluid (1-ml) was added. The test tubes were then either sealed with atmospheric gases in them or in a few cases argon or nitrogen gas was used to purge the system. The tubes were sealed by heating them toward the top while pulling on the rim with pliers. As the glass begins to soften the rim will pull away causing the section that is being heated to close which creates the seal. The tubes and contents were then heated to various temperatures (500F, 550F, 600F, and 650F) for 24-hours. They were then removed from the oven, allowed to cool to room temperature and opened. At this point, the fluid was removed and saved in vials and the coupons were washed with Freon in an ultra-sonic bath. Once clean, the coupon was re-weighed so that a weight change could be calculated.

Some of the samples were heated in vials so that GC-FTIR work could be performed on the headspace. GC-MS was also performed on the headspace but no conclusive data has been obtained as of yet. Mossbauer spectroscopy was performed on 1018 steel coupons heated in Krytox to determine the presence of FeF₃ as well as FeF₂.

RESULTS: The degradation products of Krytox, Aflunox and Demnum were analyzed with the use of GC-FTIR. The conditions in which the PFPAE fluids and the 1018 steel coupons were exposed was 600F for 24-hours. Looking at figure-1, which shows the total response chromatogram (TRC) for Krytox, one see's three large peaks which are identified from their infrared spectra shown in figures-2, 3, and 4 respectively. The first peak is identified as CO₂ while the second peak is shown by the library search to have a good match with Trifluoroacetic acid (see figure-3). The third peak looked at from figure-1 appears to belong to SiF₄ which has a IR peak located at 1028 cm⁻¹ compared to a literature value of 1030 cm⁻¹. Also looked at was a Krytox sample exposed to the same conditions but without a metal coupon. The TRC for this is shown in figure-5 which shows only the presence of CO₂. This leads one to believe that degradation of Krytox takes place at a slower rate when the metal 1018 steel is not present which hints at the idea that the metal is a catalyst. The headspace for Demnum was also looked at both with and without 1018 steel present. The TRC's obtained respectively are shown in figures 6 and 7. After a peak for peak comparison it was found that degradation products were relatively identical for both and mainly consisted of CO₂ and water. It should be noted at this point that it was discovered that SiF₄ was bleeding out of the column on this run and producing a peak around a time of seven minutes. Finally Aflunox was looked at but only with

a metal (1018 steel) present. The TRC is found in figure-8 and shows CO₂ to be present at a time of around one minute. Also found were components of degradation that have not been identified and are further being studied.

It was also a concern of ours to look into the metal-fluid reactivity to determine which fluid actually had a higher degree of interaction with the 1018 steel coupon. By just looking at the picture displaying the results for numbers 15 thru 19 (figure-9) one may not have a good idea of the reactivity but once table-1 is studied there should be little doubt about the most reactive fluid which is Perfluoropolydioxolene with a weight change of -0.45 mg/cm². The other fluids ranked in decreasing reactivity had weight changes as follows: Krytox (-0.42mg/cm²), Perfluoropolyethylene oxide (+1.41mg/cm²), Demnum (+0.81mg/cm²) and Aflunox (+0.15mg/cm²).

X-ray photoelectron spectroscopy (XPS) was used to help determine the surface composition of the metal coupons after exposure to various conditions. Looking at figure-10 one can see the oxide formation as well as the organic's present. At 500F plus air with no fluid an oxide is present and organic is not. At 550F plus air in Krytox organics appear and the oxide is still present but when the air was exchanged for nitrogen the organic presence decreased and the oxide presence disappeared. Figure-11 shows that when air is replaced with argon and Krytox is added that

fluorinated compounds begin to deposit in large quantities on the coupon's surface.

The Mossbauer spectroscopy studies revealed the presence of FeF_3 as well as FeF_2 on the surface of the coupons after being heated to 450F in Krytox for 24-hours. This data is presented in figure-12. When looking at figure-13 one sees that as the temperature was increased the reactivity with iron in Krytox increased. Also observed was that at lower temperatures (350F-550F) the reactivity was greater in the inert environment while at 600F the reactivity was nearly equal.

DISCUSSION: The results obtained through GC-FTIR at the present moment reveal very little of the degradation process but with further studies could prove to be ir aluable. We were able to see in the case of Krytox that having a metal present resulted in a quicker degradation process compared to that of without. Also determined was that, in the cases of Demnum the presence of a metal made little difference in the degradation process. Since only one case for Afflunox was looked at there is much more to be done with this fluid.

The reactivity of the different PFPAE fluids were evaluated and the end results displayed in the picture of figure-9 as well as the weight change results in table-1. It was determined that since perfluoropolydioxolene had such a large weight change ($-.045\text{mg}/\text{cm}^2$) and that large flakes of corrosion observed floating in the fluid that the reactivity must be very high. On the other hand, Aflunox had a positive weight change but very small ($+0.15\text{mg}/\text{cm}^2$). This showed the presence of corrosion as will as a metal-fluid reaction although both were small. In conclusion, the Perfluoropolydioxolene fluid was the most reactive while Aflunox the least of the PFPAE fluids looked at. Also found thru several other weight change experiments was that, of the three metals used (1018, M-50 and M-50NIL) the PFPAE fluids were more reactive towards 1018 steel. This may result from the fact that 1018 has a larger percentage of iron composition then that of M-50 or M-50NIL.

Both the XPS and Mossbauer data obtained shows signs of degradation. Since we know the composition of Krytox to contain carbon, oxygen, and fluorine we can reason that upon heating that deposits of oxides as well as fluorinated compounds will form on the surface of the metal.

The conclusions drawn from this search to date is the following: 1.) 1018 steel is much more reactive than M-50 and M-50NIL. 2.) Perfluoropolydioxolane is the most reactive fluid in the presence of 1018 steel and Aflunox the least. 3.) Reactivity is strongly dependant on the fluid-metal pair. And 4.) There is a significant difference in fluid-metal reactivity depending on the test environment and test duration.

FIG 1

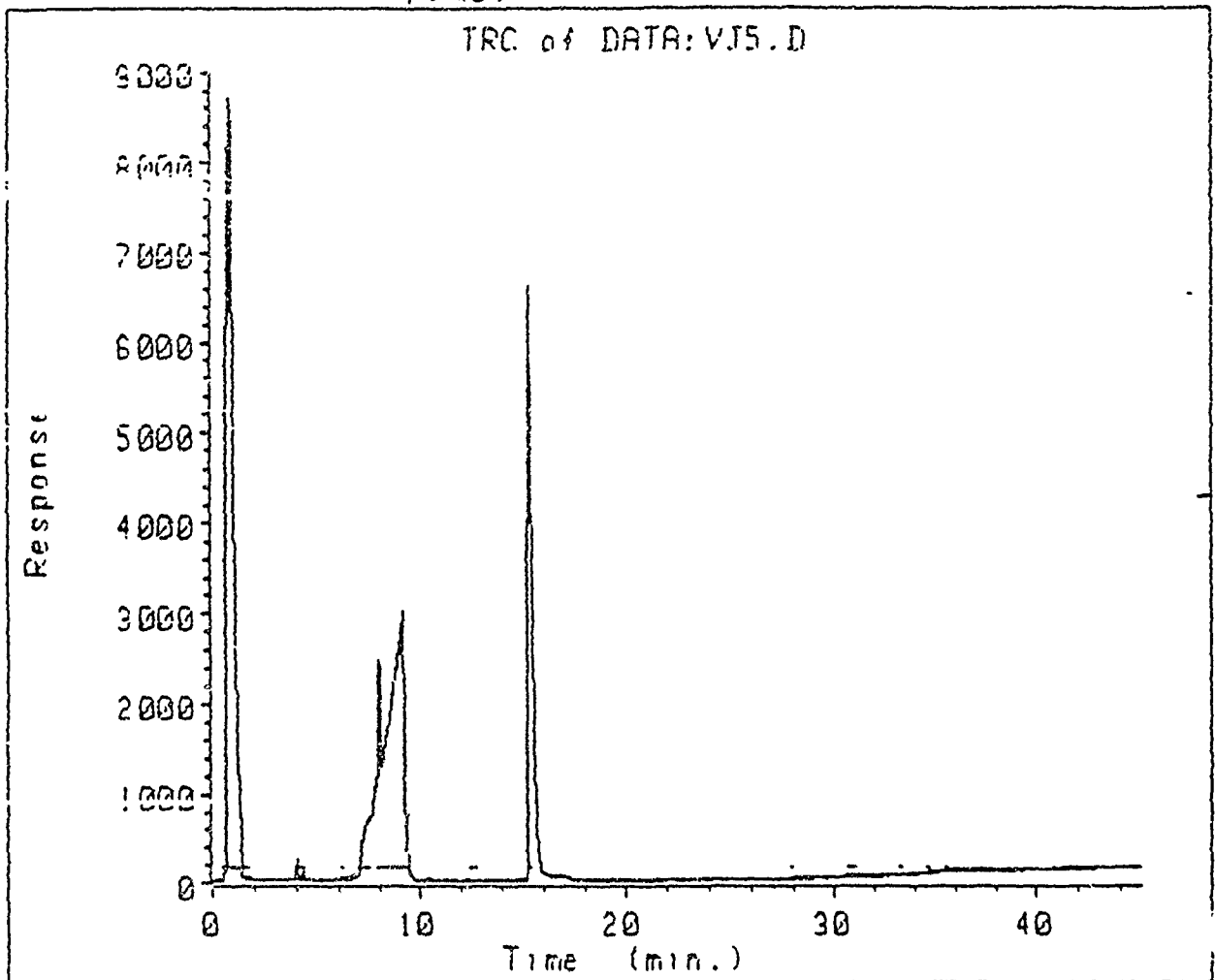


FIGURE-1

Total Response Chromatogram for Krytox in the presence of
1018 steel heated to 600F for 24-hrs

FIG 82

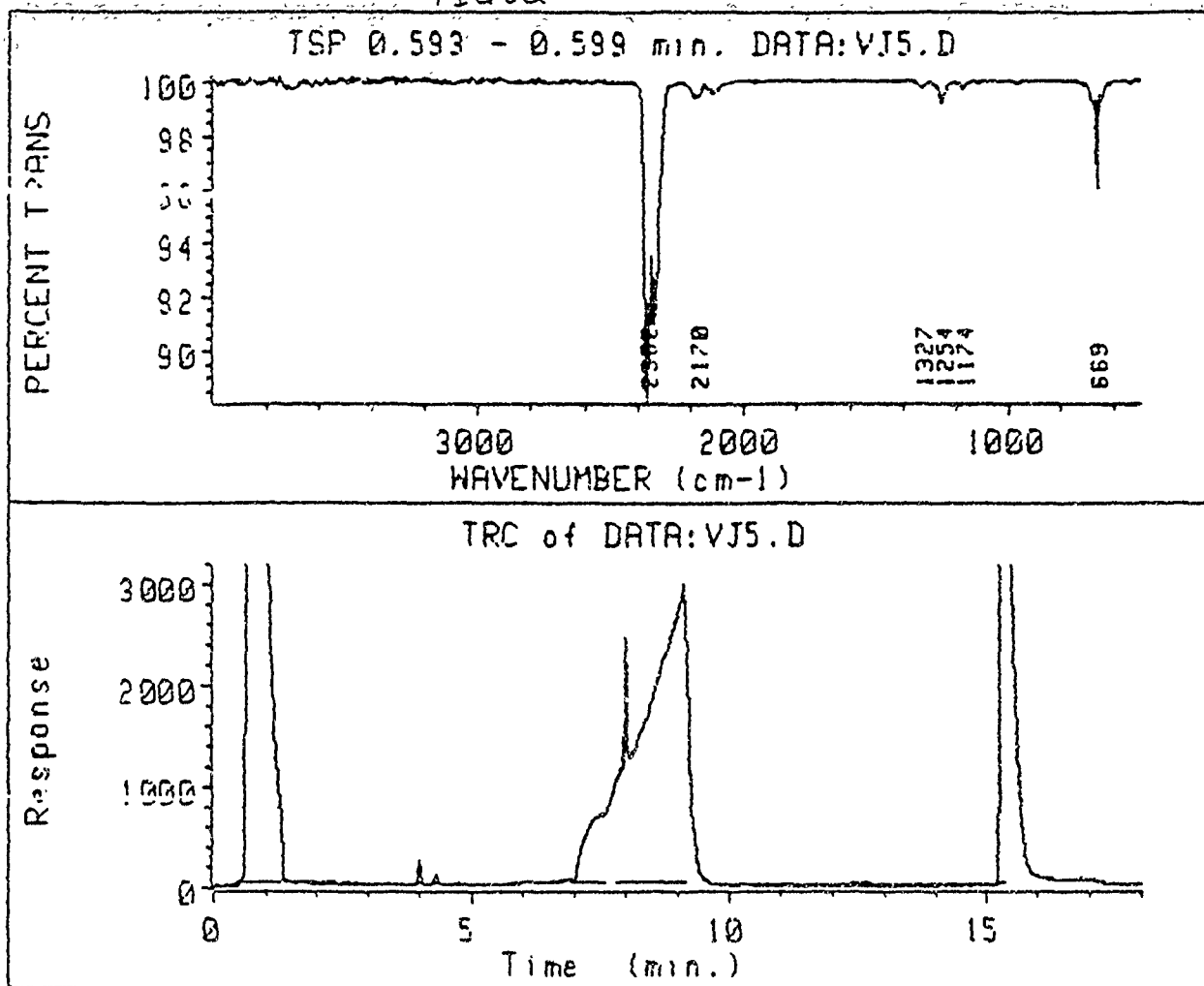


FIGURE-2

IR Spectra of Figure-1 for the time interval of 0.593-0.599 min.

FIG 82

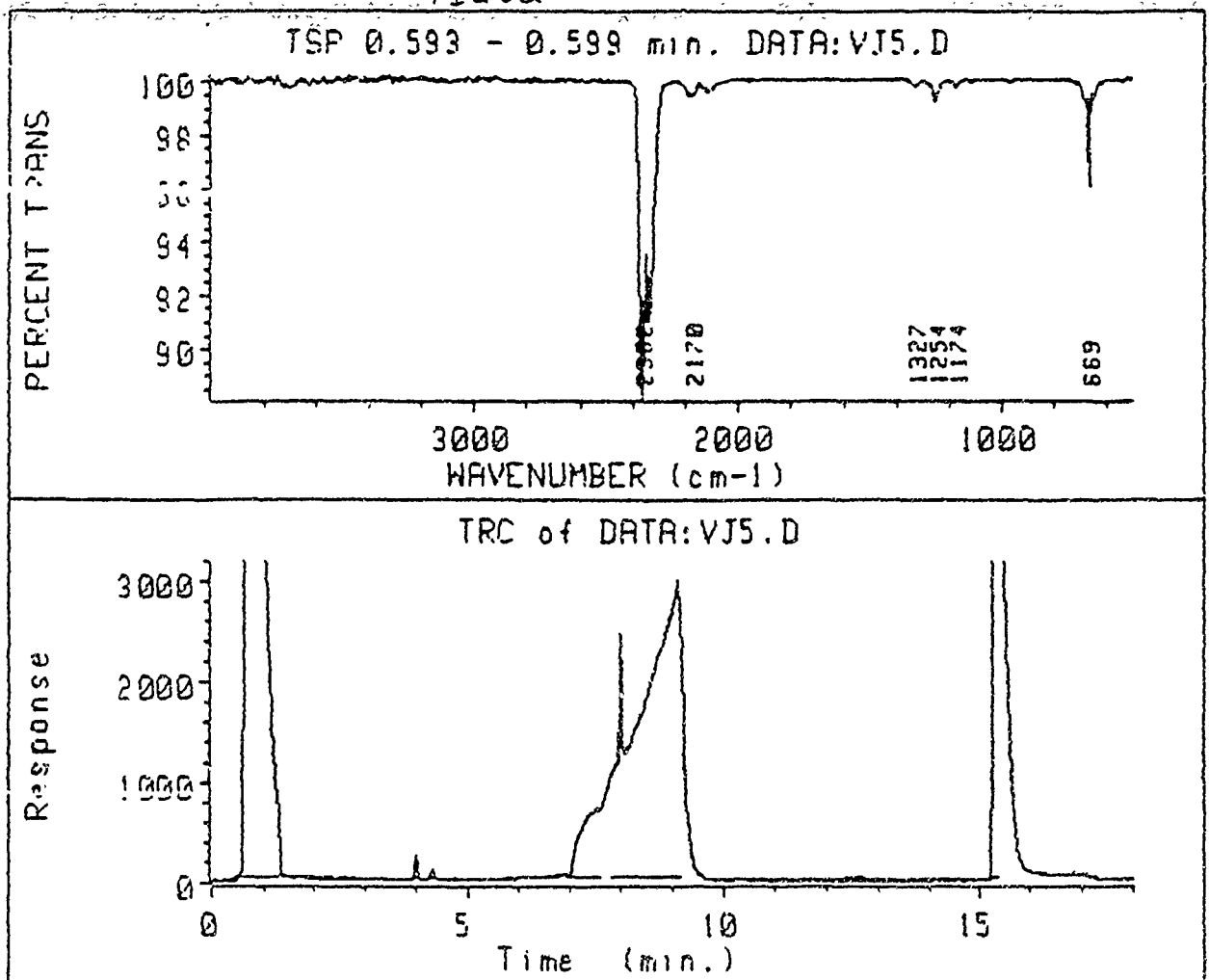
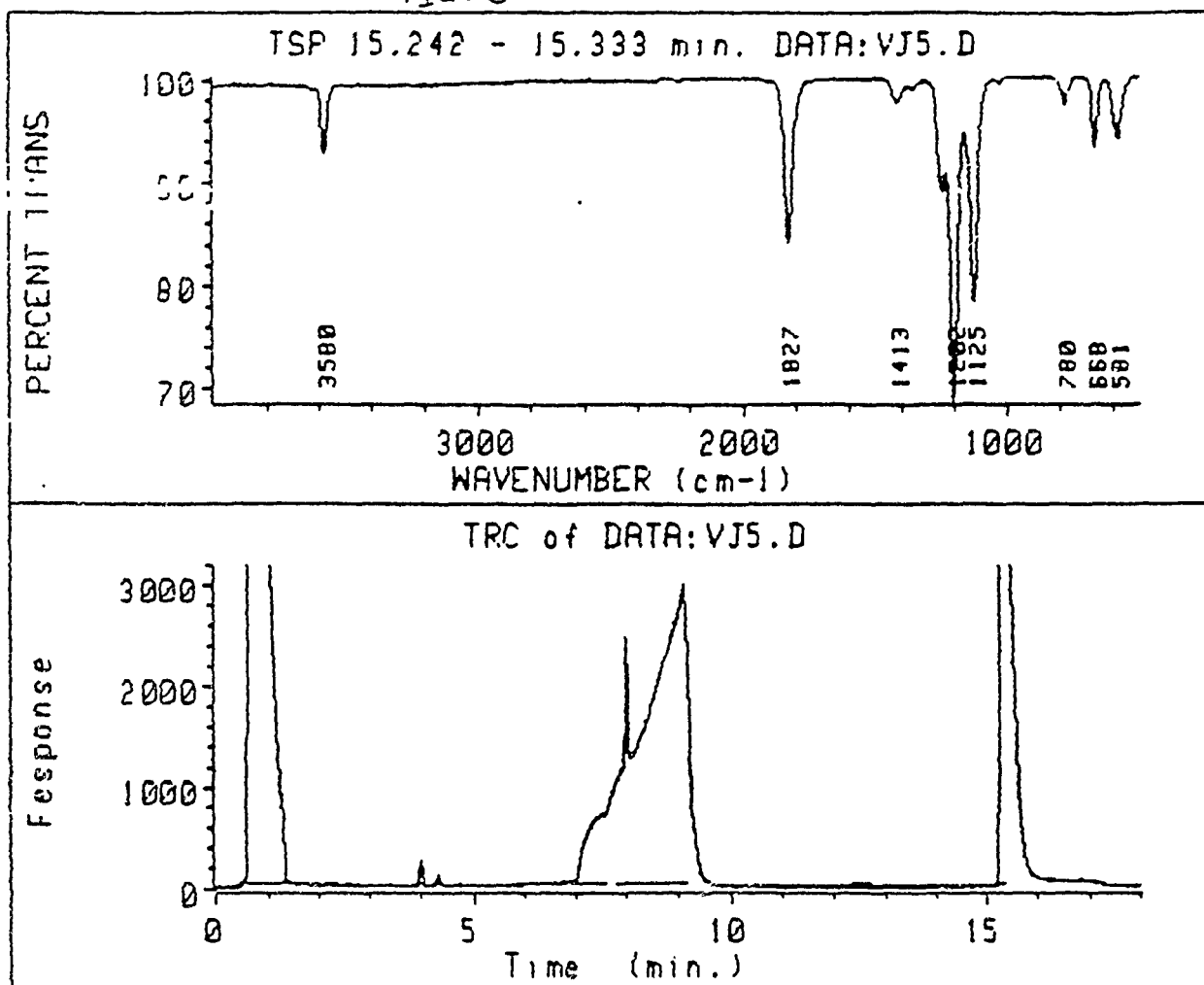


FIGURE-2

IR Spectra of Figure-1 for the time interval of 0.593-0.599 min.

FIG 83



GRAPHICS RESULTS

Version 1.0 25-Feb-87

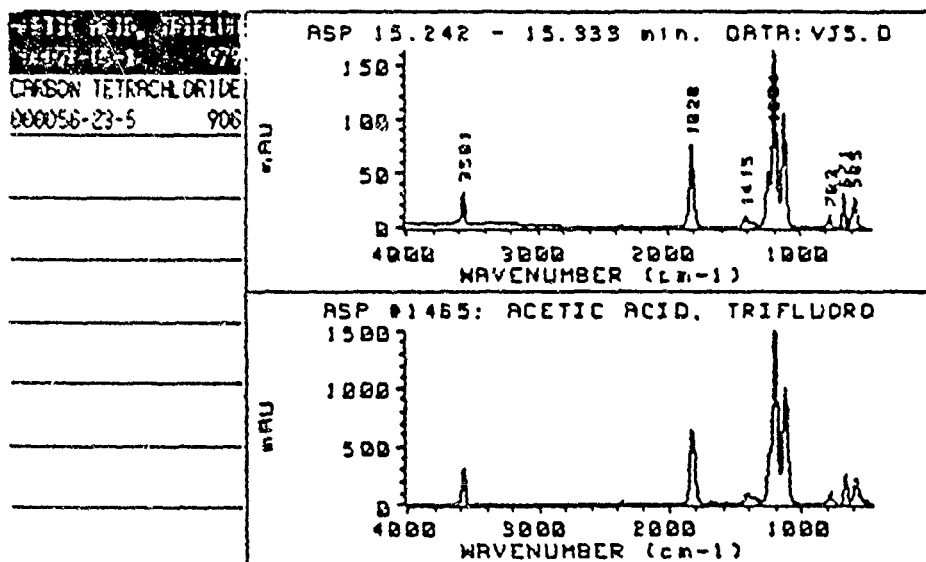


FIGURE-3

IR Spectra of figure-1 for the time interval of 15.242-15.333 min.

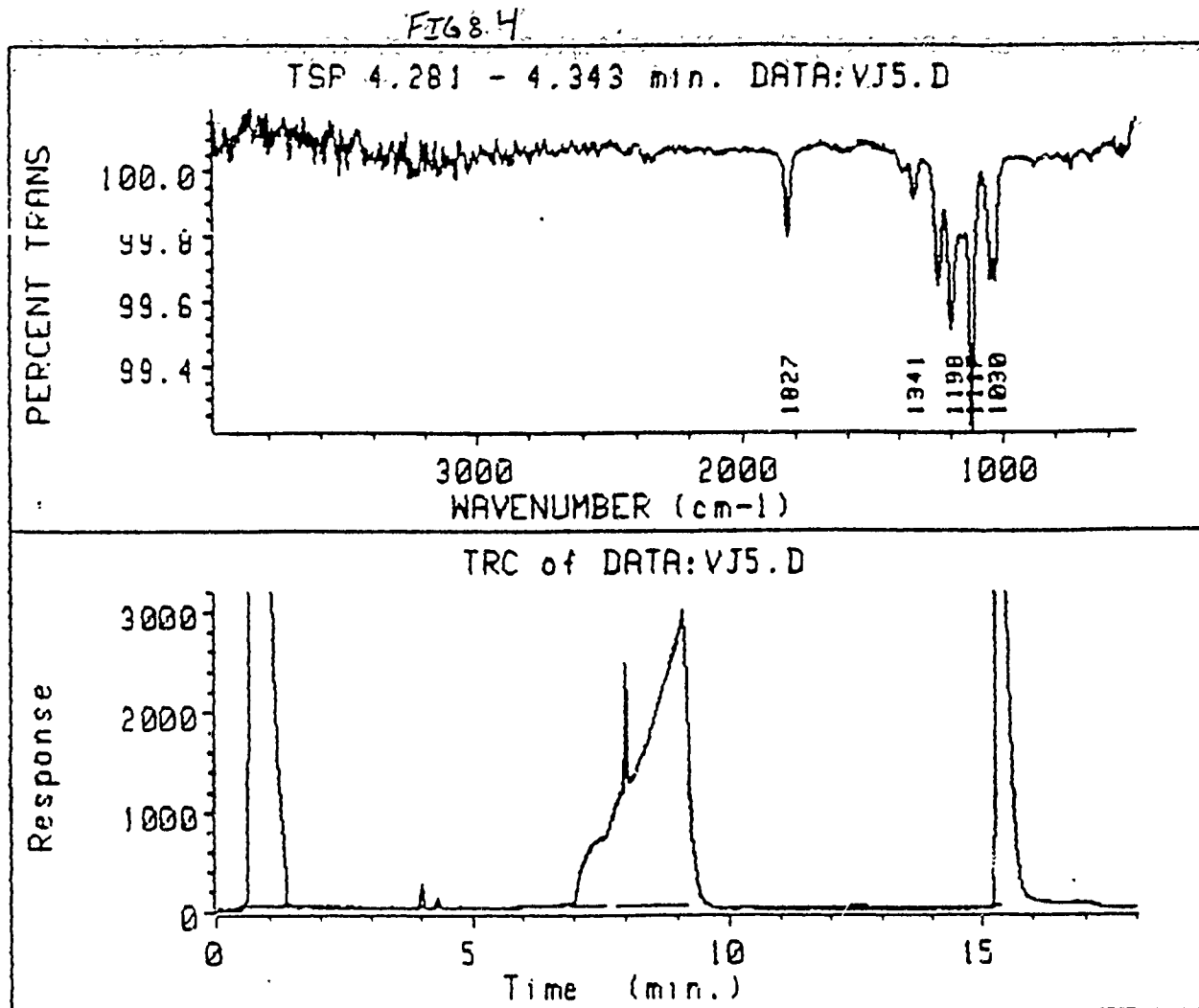


FIGURE-4

IR Spectra of Figure-1 for the time interval of 4.281-4.343 min.

FIG:5

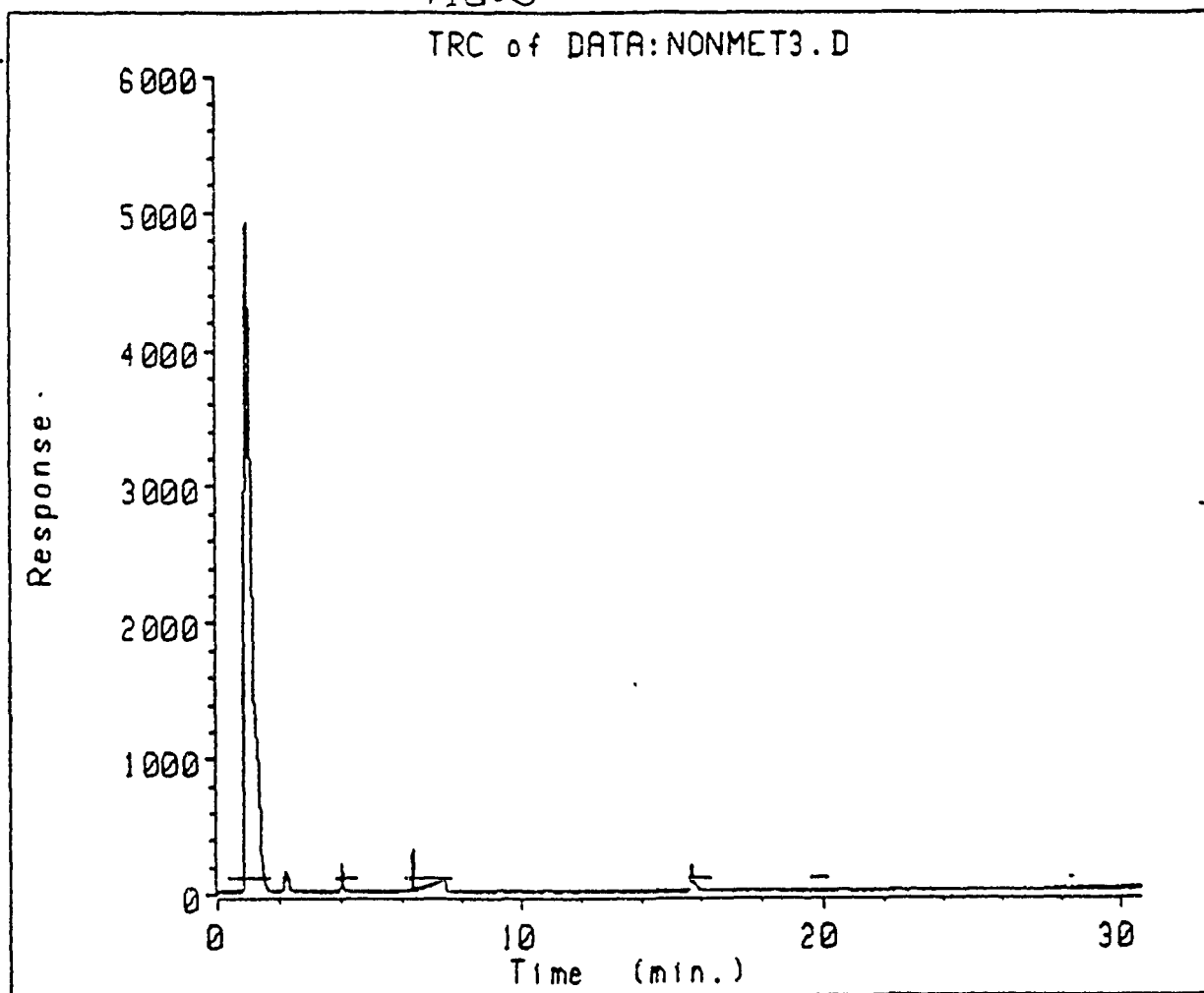


FIGURE-5

Total Response Chromatogram for Krytox without metal coupon heated to 600F for 24-hrs.

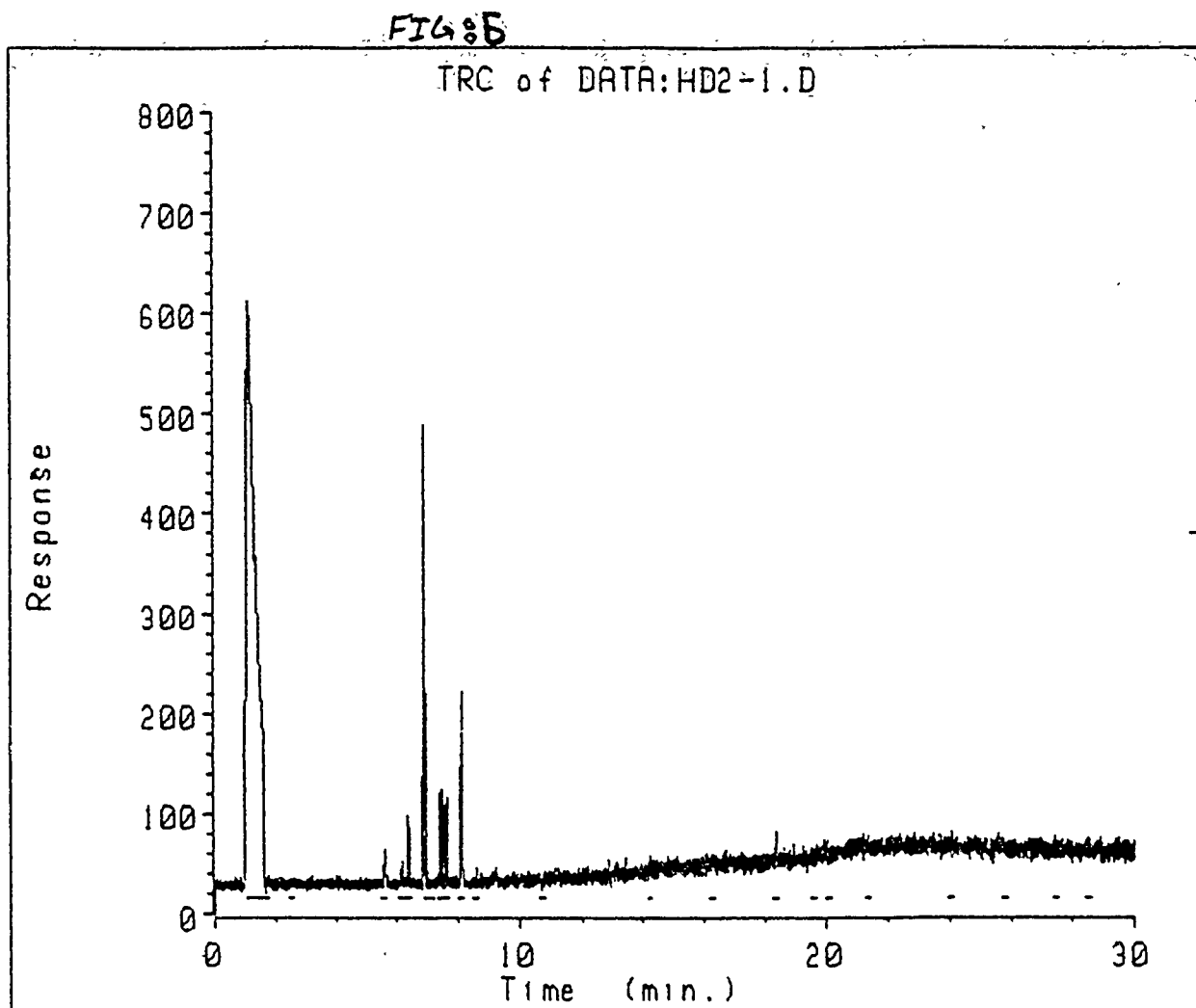


FIGURE-6

Total Response Chromatogram for Demnum in the presence of 1018 steel heated to 600F for 24-hrs.

FIG 87

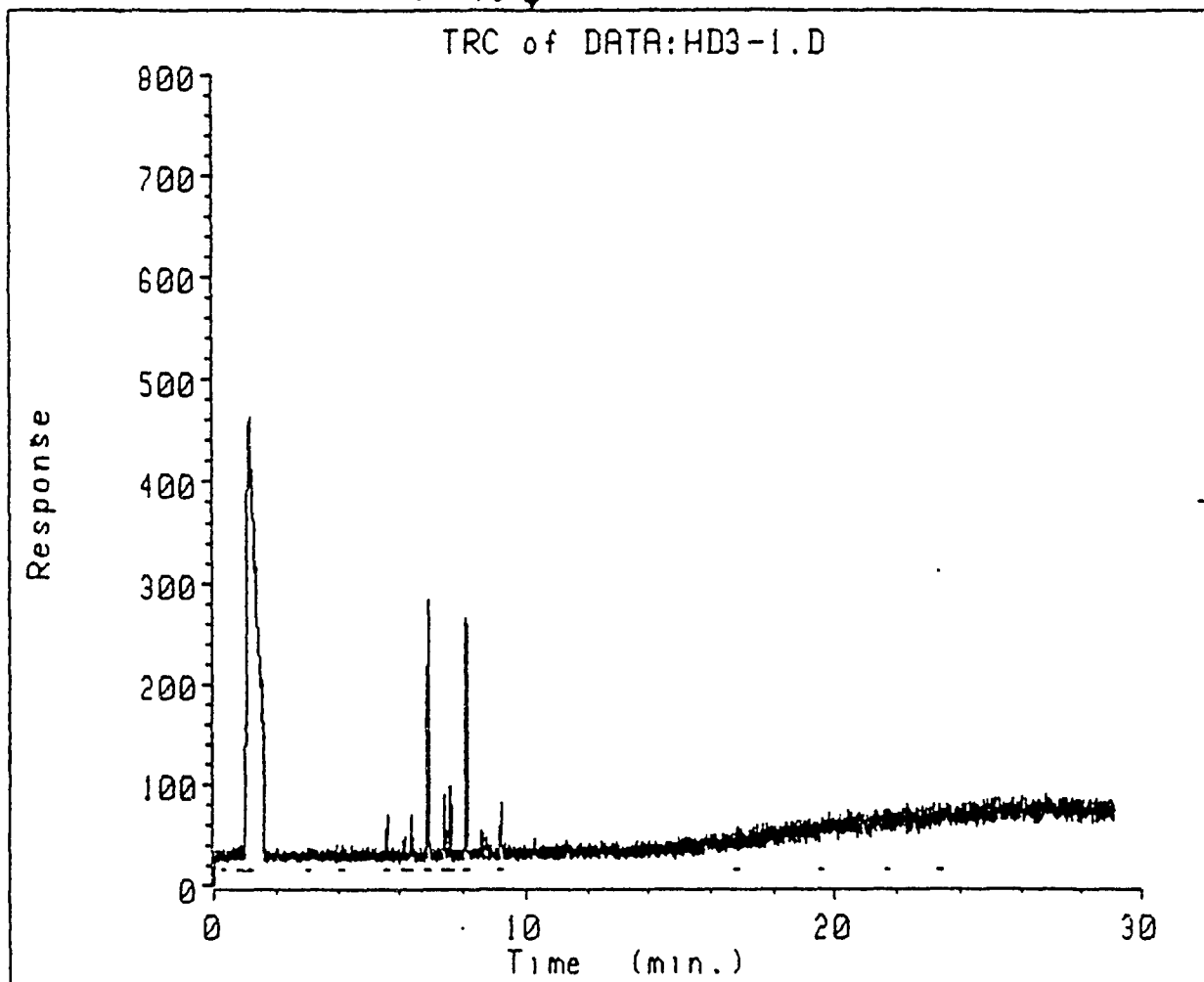


FIGURE-7

Total Response Chromatogram for Demnum without a metal coupon heated to 600F for 24-hrs.

FIG: 8

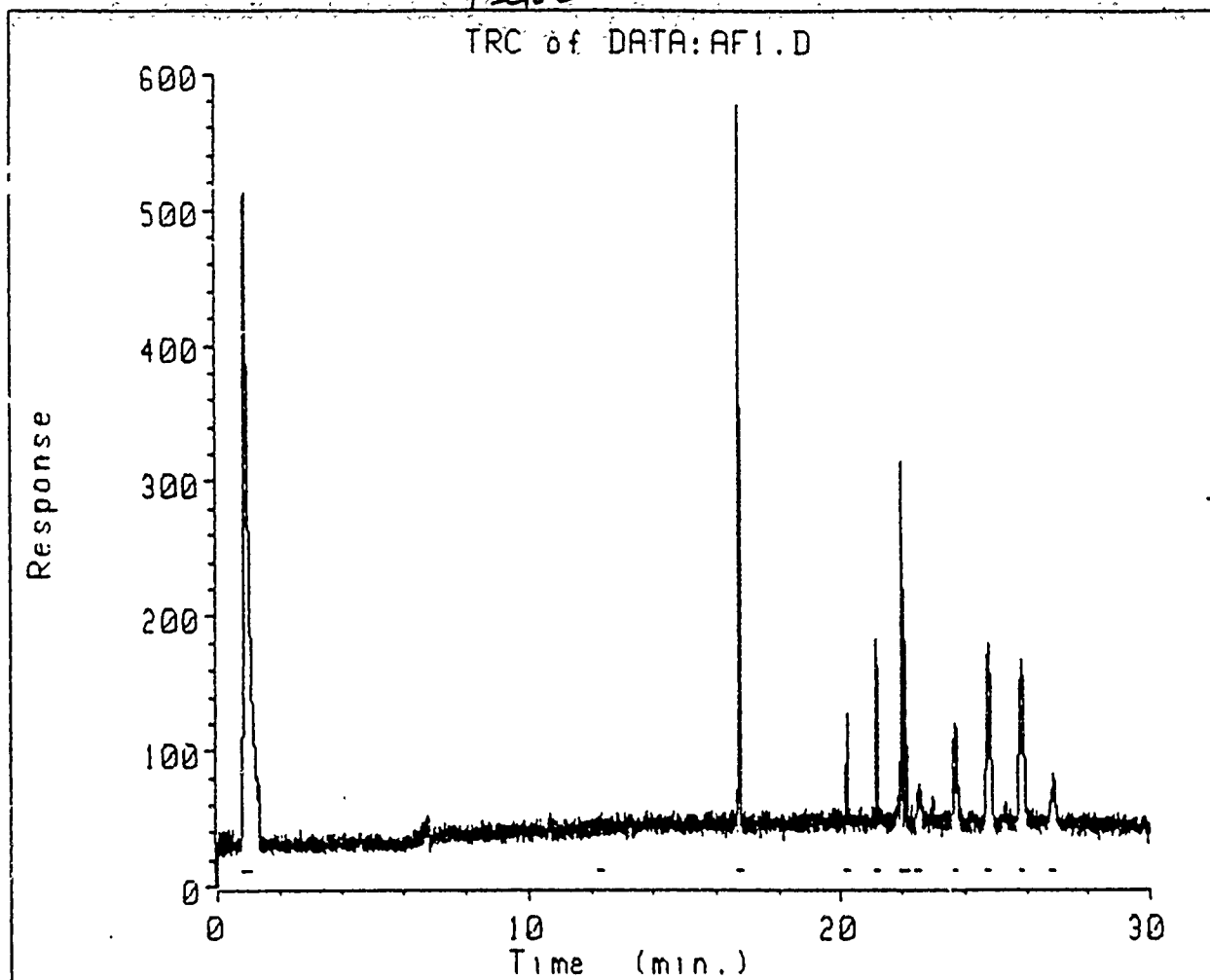


FIGURE-8

Total Response Chromatogram for Aflunox in the presence of 1018 steel heated to 600F for 24-hrs.

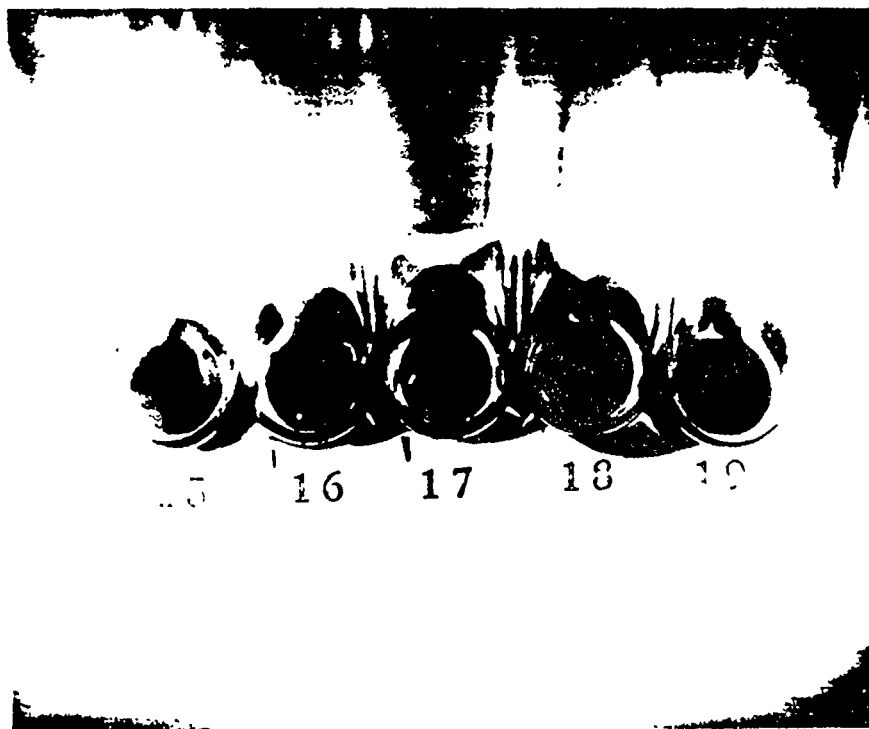


FIGURE-9

TABLE-1

PICTURE	PFPPE	WT. CHANGE (mg/cm ²)
#15	Perfluoropolyethylene oxide	+1.41
#16	Perfluoropolydioxolene	-0.45
#17	Afflunox	+0.15
#18	Demnum	+0.81
#19	Krytox	-0.42

CONDITIONS: 550F, 24-hours, 1018 steel coupon

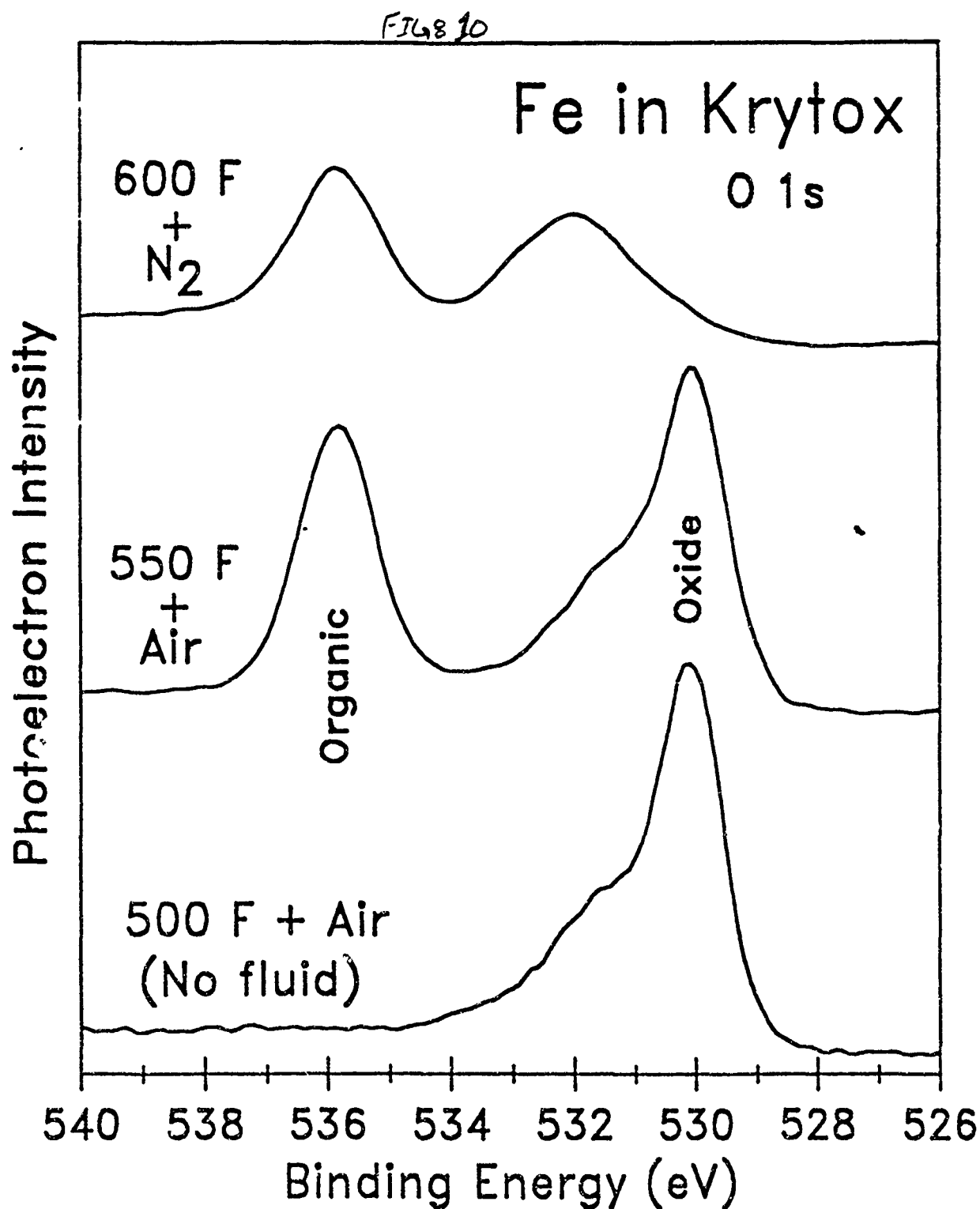


FIGURE-10

XPS data for oxide presence under different environments as a function of temperature.

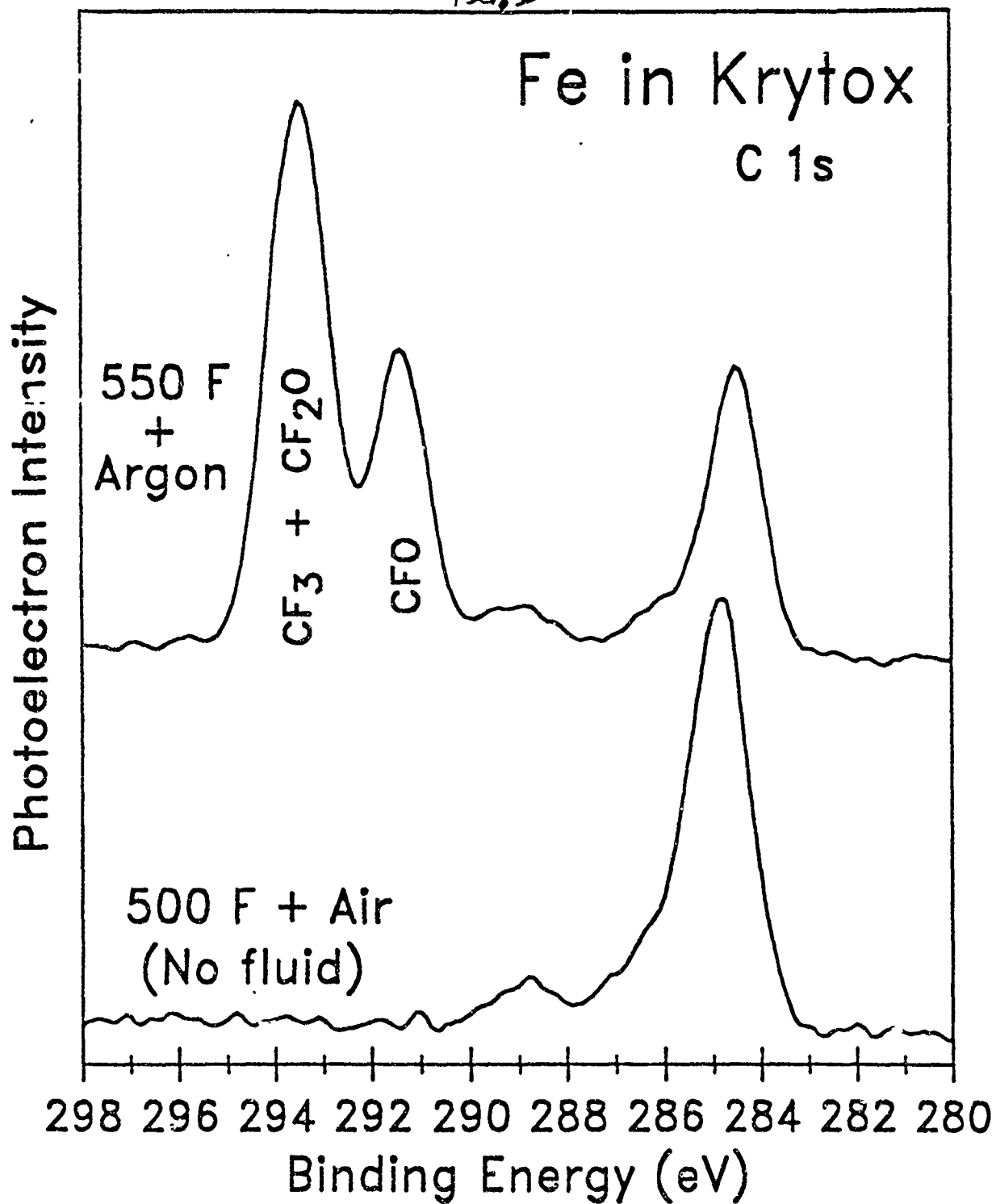


FIGURE-11

XPS data for fluorinated compound presence under different environments as a function of temperature.

Fig 812

1018 Steel - Krytox AC, 450°F - 24h

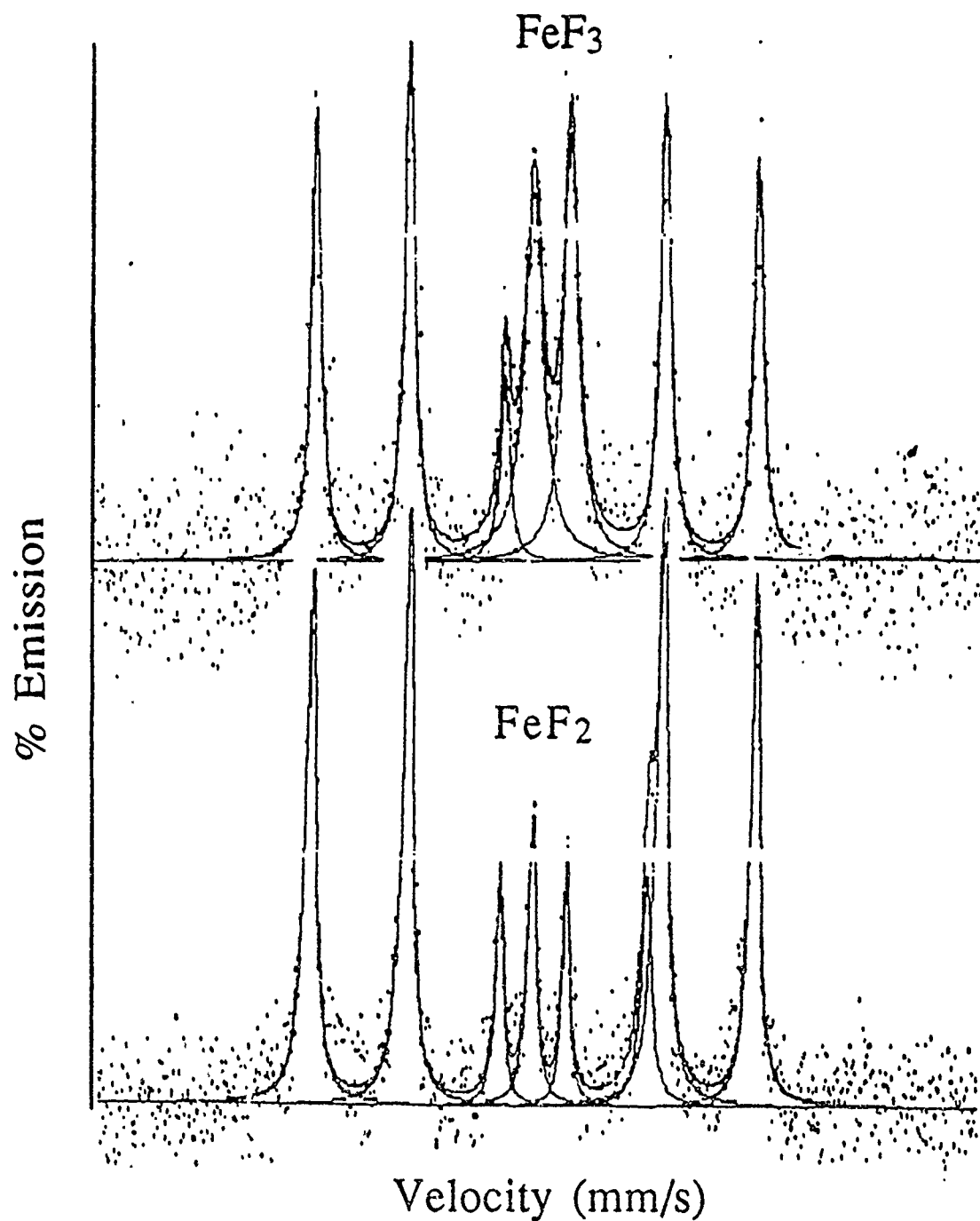


FIGURE-12

XPS data for the presence of FeF_3 and FeF_2 .

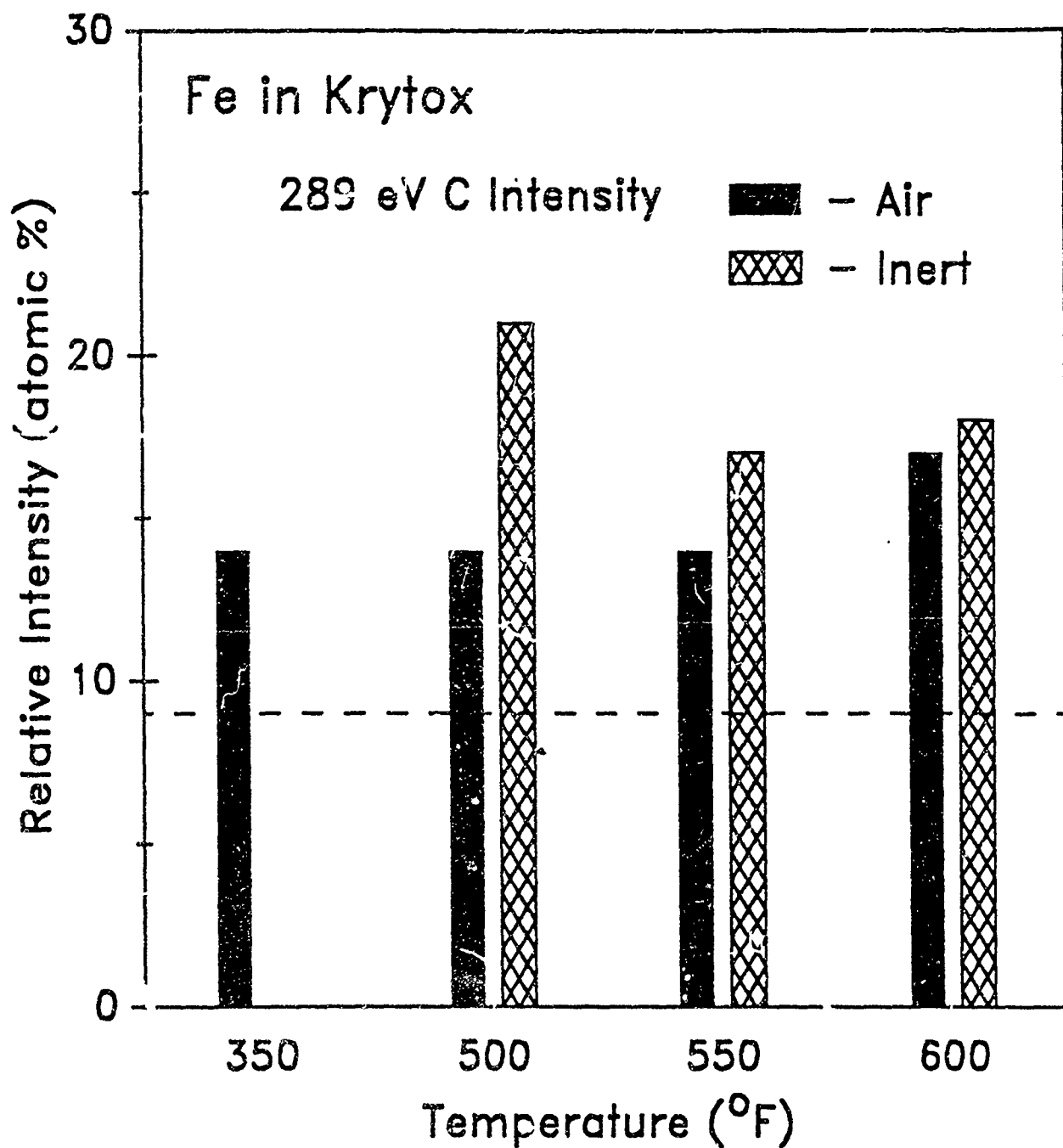


FIGURE-13

XPS data for Fe in Krytox under different environments as a function of temperature.

REFERENCES

1. Snyder, C.E., Gschwender, L.J., "PFPAE-Liquid Lubricants of the Future"
2. Snyder, C.E., Dolle, R.E., "Development of Perfluoroalkylethers as High Temperature Lubricants and Hydraulic Fluids," vol 19, 3, 171-180
3. Carre, D.J., "Perfluoropolyalkylether Oil Degradation: Inference of FeF₃ Formation on Steel Surfaces under Boundary Conditions," vol 29, 2, 121-125
4. Liang, J., July 91 monthly report.

NUMERICAL SIMULATION OF PHOTOREFLECTANCE FROM TWO-DIMENSIONAL ELECTRON GAS IN A GaAs/AlGaAs HETEROJUNCTION

by

James R. Engholm

Physics Dept.

University of Minnesota, Duluth

RDL Summer Graduate Student Research Program

ID#301, for Dr. William Mitchel

at MLPO/WPAFB

Summer 1991

ABSTRACT: We present a first principles calculation of the photorefectance lineshape originating from the quantum well region of a GaAs/Al_xGa_{1-x}As heterojunction.

INTRODUCTION:

A typical High Electron Mobility Transistor (HEMT) heterojunction structure¹ (Figure 1) consists of first a 400nm surface layer of doped Al_xGa_{1-x}As (we used $x=3$) with a donor dopant density (N_d) of $2 \times 10^{18} \text{cm}^{-3}$. Below this is a 10nm undoped Al_xGa_{1-x}As spacer layer, under which is a 500nm layer of GaAs which typically has an unintentional impurity acceptor density (N_a) on the order of $1 \times 10^{14} \text{cm}^{-3}$. The interface below the spacer layer is referred to as a heterojunction. The band structure associated with these structures is shown in Figure 2, where the conduction band interface offset used is 300meV.

Note that the conduction band forms a quantum well structure just below the heterojunction. This well restricts the z-direction movement of any trapped electrons. Thus the electrons form a two-dimensional electron gas (2DEG) in a plane parallel to the sample surface.

Photoreflectance² (PR) is a method of studying the electronic band structure of semiconductors. It relies on detecting the change in the amplitude of the light reflected from the near-surface regions of the sample caused by the laser-induced periodic modulation of the band structure. While it illuminates the sample, the laser promotes electrons from the valence band to the conduction band leaving holes in the valence band. If there is an electric field present (ie., if the bands are bent), as is true in many near surface semiconductor structures, the electrons and holes created will alter the local structure of the band, thus affecting the local dielectric function, which determines the reflective properties of the sample. The laser is modulated by chopping the laser beam - ie. by periodically blocking and passing the beam (a 200Hz chopping frequency is typical).

In our model, we follow the method of Snow et. al.³, who modeled the PR signal originating from the heterostructure by assuming that the modulation mechanism is the variation in the 2DEG density in the well due to laser-promoted electrons. We used an iterative, self-consistent method to determine the conduction band structure and the electron envelope wavefunctions in the well for both the laser-illuminated and unilluminated cases, assuming that equilibrium was reached in both cases. These wavefunctions and the

calculated wavefunctions for the valence band states were used to calculate the change in the photon induced transition rate, from which the change in the absorption profile was deduced. The changes in the dielectric function are calculated from the absorption changes, and from these the photoreflectance lineshape can be determined. In all steps, we calculated only the behavior of the quantity involved, not the actual magnitude.

DISCUSSION OF METHOD:

DETERMINATION OF BAND STRUCTURE: Given the donor and acceptor densities, the spacer width, the donor binding energy (we used 48meV), the dielectric constant of GaAs (11.56), and the effective electron masses in the two materials ($.088m_0$ in $Al_xGa_{1-x}As$ and $.0665m_0$ in GaAs, where m_0 is the electron rest mass), we calculated the conduction band structure, the confined wavefunctions for the first two bound states of the well, and the 2DEG density (N_s). To this end we adapted a numerical routine created by Thomas A. Vaughan and Dr. Frank Szmulowicz of UDRI/MLPO/WPAFB

Once the conduction band structure is determined, the valence band structure is calculated by simply subtracting the energy gap in each region (1.8eV for $Al_{.3}Ga_{.7}As$ and 1.424eV for GaAs). The valence band electron envelope wavefunctions can then be calculated with the same finite-difference method discussed above. A typical calculated band structure is shown in Figure 2, where $E=0$ corresponds to the bottom of the quantum well at $z=0$.

TRANSITION RATE CALCULATION: The transition rate (w) from an unbound valence band state v to a bound conduction band state c is given by^{4,5}

$$w_c(h\nu) \propto \int_{E_v=E_c-h\nu}^{E_{v0}} \frac{| \langle v | c \rangle |^2}{\sqrt{E_v - E_{v0}}} dE_v \quad (1)$$

which is just an integral over the valence band states accessible by a given photon energy ($h\nu$) of the overlap integral squared of valence band and conduction band envelope wavefunctions, multiplied by the density of valence energy states.

ABSORPTION CALCULATION AND MODULATION: The absorption (α) as a function of photon energy behaves as the transition rate divided by the photon energy (α goes as $w/h\nu$)⁵.

We proceeded through the above calculation leading to the absorption twice for each data run - once for the unilluminated case and then once for the illuminated case, for which we added a small quantity to the 2DEG concentration. Each absorption calculation included heavy hole and light hole contributions (they are degenerate in energy at the fundamental band edge of GaAs). Thus we had two absorption curves as functions of photon energy, which we subtracted pointwise to get the change in the absorption due to modulation $\Delta\alpha = \alpha_{\text{unillum}} - \alpha_{\text{illum}}$.

Two typical absorption curves are shown in Figure 3. One is the unilluminated case, and the other is an illuminated case with the 2DEG modulation set to $\delta N_s = 5 \times 10^{10} \text{ cm}^{-2}$.

DIELECTRIC FUNCTION MODULATION: The complex dielectric function of a material can be expressed as $\epsilon = \epsilon_1 + i\epsilon_2$. Similarly the

change in the dielectric function due to modulation can be expressed $\Delta\epsilon = \Delta\epsilon_1 + i\Delta\epsilon_2$. In general, ϵ_2 is related to the absorptive qualities of the material, while ϵ_1 is related to the reflective/transmissive qualities. $\Delta\epsilon_2$ can be related pointwise to the change in the absorption by $\Delta\epsilon_2$ goes as $\Delta\alpha/h\nu^5$.

$\Delta\epsilon_1$ can be derived from the knowledge of $\Delta\epsilon_2$ over all energies through a Kramers-Kronig (K-K) integral transformation⁶

$$\Delta\epsilon_1(\omega) = \frac{2}{\pi} \int_0^\infty \frac{s \Delta\epsilon_2(s)}{s^2 - \omega^2} ds \quad (2)$$

where ω is the frequency of the photon (which could just as easily be $h\nu$, the energy). This integral equation states that the shape of $\Delta\epsilon_1$ at a given photon energy depends most heavily on the shape of $\Delta\epsilon_2$ at the same energy, since this is where the denominator is smallest, while the denominator reduces the contributions at photon energies with greater differences in energy.

PHOTOREFLECTANCE CALCULATION: PR can be related to the changes in the dielectric function by²

$$\frac{\Delta R}{R}(h\nu) = \alpha(h\nu) \Delta\epsilon_1(h\nu) + \beta(h\nu) \Delta\epsilon_2(h\nu) \quad (3)$$

where α and β are called the Seraphin coefficients⁷ (the former not to be confused with absorption). These coefficients vary with energy, and their behavior has been approximated by several short lines for our purposes. Note that frequently in PR analysis, β is assumed to be small enough to ignore, but we have included it here.

Figure 4 shows the calculated PR for the case of $\delta N_s = 5 \times 10^{10} \text{cm}^{-2}$.

RESULTS:

For our typical values of $N_d=2 \times 10^{18} \text{cm}^{-3}$ and $N_a=1 \times 10^{14} \text{cm}^{-3}$, our results were $E_1=60.0 \text{meV}$ and $E_2=84.3 \text{meV}$ above the bottom of the well (which we defined as $E=0$). Figure 2 shows the band structure plotted against depth, where $z=0$ is defined as the heterojunction. These parameters led to a ground state 2DEG density of about $9.24 \times 10^{11} \text{cm}^{-2}$ with a Fermi energy of $E_F=76.5 \text{meV}$ above the bottom of the well. Thus the first excited state is unoccupied since we have assumed zero temperature.

Upon adding a modulation 2DEG density of $\delta N_s=5 \times 10^{10} \text{cm}^{-2}$, the results change to $E_1=61.8 \text{meV}$ and $E_2=86.7 \text{meV}$ while the Fermi energy rises to 79.4meV .

CALCULATED PHOTOREFLECTANCE SPECTRA: Our 2DEG density modulation PR data came to be grouped into two broad categories: (a) large amplitude modulation, and (b) small amplitude modulation.

(a) Figure 4 shows the PR with $\delta N_s=5 \times 10^{10} \text{cm}^{-2}$ and $N_a=1 \times 10^{14} \text{cm}^{-3}$. This PR is a typical large amplitude modulation. It has broad peaks at 1.46eV and 1.52eV , along with smaller peaks at energies near the band gap. The typical large amplitude modulation structure has 2-4 broad peaks between just above the band gap to 1.55eV , and possibly several small, sharp oscillations near the band gap.

The origin of the two broad peaks in Figure 4 can be seen from the absorption curves of Figure 3. There is clearly a relatively large amount of absorption modulation in the flat region of the curves around 1.46eV and in the second rising region, found from 1.49eV to 1.52eV , which give rise to the broad peaks. The secondary, smaller

peaks at lower energy come from the smaller modulation of the initial steep rising edge of the absorption curve near the band gap.

(b) Figure 5 shows the PR spectra with $\delta N_s = 1 \times 10^6 \text{ cm}^{-2}$ and $N_a = 1 \times 10^{14} \text{ cm}^{-3}$. This is a typical example of small amplitude modulation PR. It has a series of about 3 sharp oscillations centered around the band gap, up to about 1.45eV. Then from 1.47eV to 1.55eV it has several broader oscillations. All of these characteristics are typical of small amplitude modulation PR signals.

These sharp structures are not, we believe, a facet of a bug in the program, since when 0 modulation is requested, 0 PR is calculated by our program. In fact, they come from small modulations in the two rising edges of the absorption profile, with the sharp structures originating from the steep initial rise at and just above the band gap, and the broader, higher energy signals originating from the shallower rise around 1.50eV, as can be seen by comparing its $\Delta\alpha$ (Figure 6) with the unilluminated absorption curve of Figure 3. An explanation for the structures eludes us, however.

We explored a wide range of δN_s values, from $1 \times 10^4 \text{ cm}^{-2}$ to $1 \times 10^{12} \text{ cm}^{-2}$, while at the same time we explored a relatively broad range of N_a values, from $1 \times 10^{12} \text{ cm}^{-3}$ to $5 \times 10^{16} \text{ cm}^{-3}$, while concentrating on N_a values from $1 \times 10^{14} \text{ cm}^{-3}$ to $1 \times 10^{15} \text{ cm}^{-3}$.

Several trends in the PR data have been noticed. Firstly, the behavior with changing N_a varies with the choice of δN_s , and vice-versa. In general for combinations of smaller δN_s and smaller N_a , we see small amplitude modulation signals, and for large values of both quantities, we see large amplitude modulation signals. In fact, for a small enough values of both quantities, ie. as the well region becomes

shallower and broader, the signal tends to become similar to a typical GaAs bulk signal, typified by sharp, near band gap structures (the necessity of extremely small 2DEG modulation in this case is somewhat peculiar, and may be more a function of the well losing its 2DEG type behavior in favor of bulk behavior). For 2DEG modulation quantities above about $1 \times 10^{10} \text{cm}^{-2}$, we see no small amplitude modulation behavior at all for larger values of N_a , while below this we see a small region of small amplitude behavior between $N_a = 1 \times 10^{14} \text{cm}^{-3}$ and $N_a = 1 \times 10^{15} \text{cm}^{-3}$ for larger 2DEG modulations, along with the region of low values of N_a .

CONCLUSIONS:

COMPARISON WITH REALITY: Figure 7 shows a comparison of a calculated first excited level transition PR curve with $\delta N_s = 3 \times 10^5 \text{cm}^{-2}$ to a Differential Photoreflectance⁸ (DPR) data run of a HEMT sample. DPR is a method of suppressing the PR from any given structure in a sample. In this case, the PR due to the large surface electric fields was suppressed, leaving only the PR due to the heterojunction. This DPR data is typical of much of that acquired from heterojunction structures.

Notice also the apparent lack of any correlation between the calculated high energy, broader oscillations and the actual data. This is an intriguing observation, which means that either there is no actual modulation in the second rising edge of the absorption curve, or (anticlimactically) that they are just washed out in the noise, in which case one might argue that the oscillations between 1.45eV and 1.49eV are actually noisy data and not just noise.

Note that the DPR data set is taken at room temperature, while the calculation is for zero temperature (which implies that states above the Fermi energy are empty and states below are full). Thus a precise comparison is not advisable between the two cases.

ENERGY LEVELS: We see no reason for associating the actual bound state energies with specific heterojunction PR structures, as has been done previously. The actual energies of the levels principally affect only the gross energy shifts of an otherwise independent signal, since they are the base from which the photon energies are subtracted in the calculation. Although this shape of the well determines the energy levels, and it is the modulation of the shape which determines the PR, the bound state energy has only a cursory effect.

SUMMARY:

We have calculated from first principles the PR signal originating from transitions to the first and second energy levels of a GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ heterojunction quantum well. We observe and confirm both the broad, large modulation PR structures previously calculated, and the sharp and narrow PR structures which PR experiments indicate originate from the heterojunction. In addition, our calculations show that the PR signals originating from the heterojunction are highly dependent upon the unintentional impurity levels in the GaAs region.

REFERENCES

- ¹ See, for example, C. Kittel, Introduction to Solid State Physics (John Wiley & Sons, New York, 1986).
- ² See, for example, D.E. Aspnes, in Handbook on Semiconductors, edited by T. S. Moss (North-Holland, New York, 1980), Vol. 2, p. 109 and references cited therein.
- ³ Snow, E.E., O.J. Glembocki, and B.V. Shanabrook, Phys. Rev. B, 38, 12484, (1988).
- ⁴ See, for example, G. Bastard, Wave Mechanics Applied to Semiconductor Heterostructures (Halsted, New York, 1988).
- ⁵ Aspnes, D.E., B. Handler and D.F. Blossey, Phys. Rev., 166 No. 3, 921 (1968).
- ⁶ See, for example, M. Cardona, Modulation Spectroscopy (Academic Press, New York, 1969).
- ⁷ Seraphin, B.O., and N. Bottka, Phys. Rev., 145 No. 2, 628 (1966).
- ⁸ Sydor, M., A. Badakhshan, J.R. Engholm and D.A. Dale, Appl. Phys. Lett., 58, 948 (1991).

S.I. GaAs Substrate

Und. GaAs 1000Å	Buffer
10 30X30 GaAs/AlGaAs Q.Wells	
Und. GaAs Channel 10000 Å	2DEG
Und. AlGaAs Spacer	
2.0E18 Si AlGaAs X=0.3	
GaAs Cap	

Figure 1

Figure 2

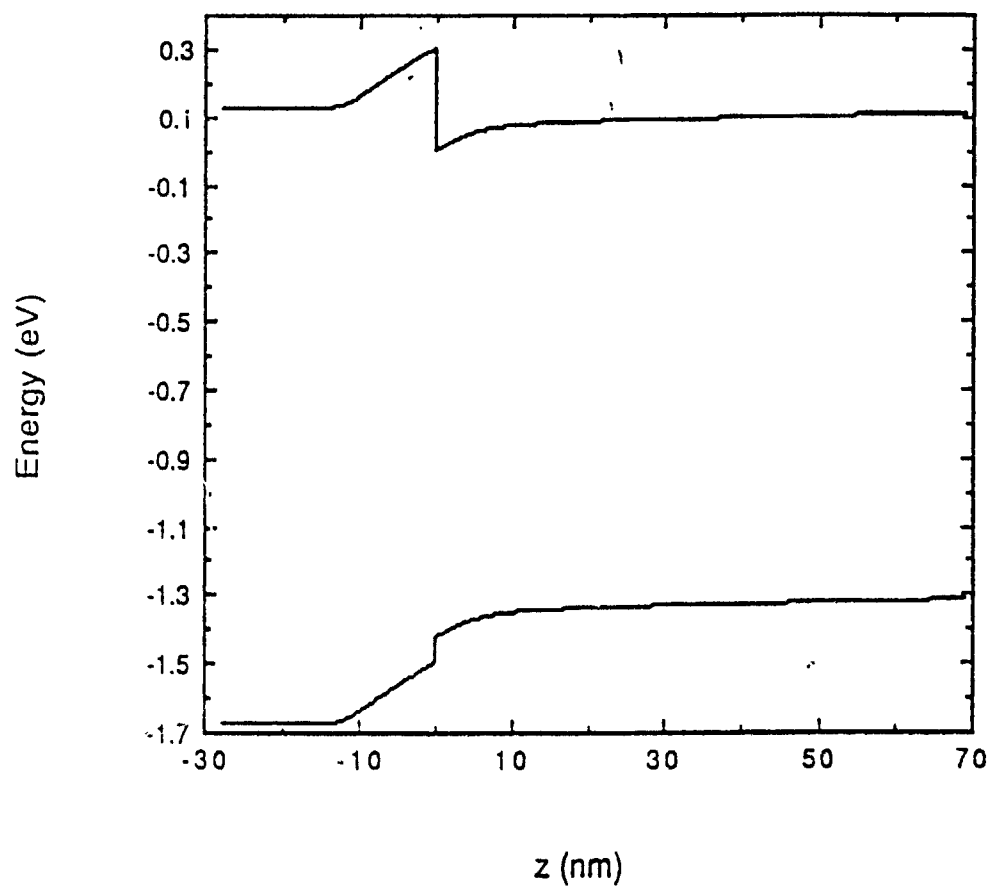


Figure 3

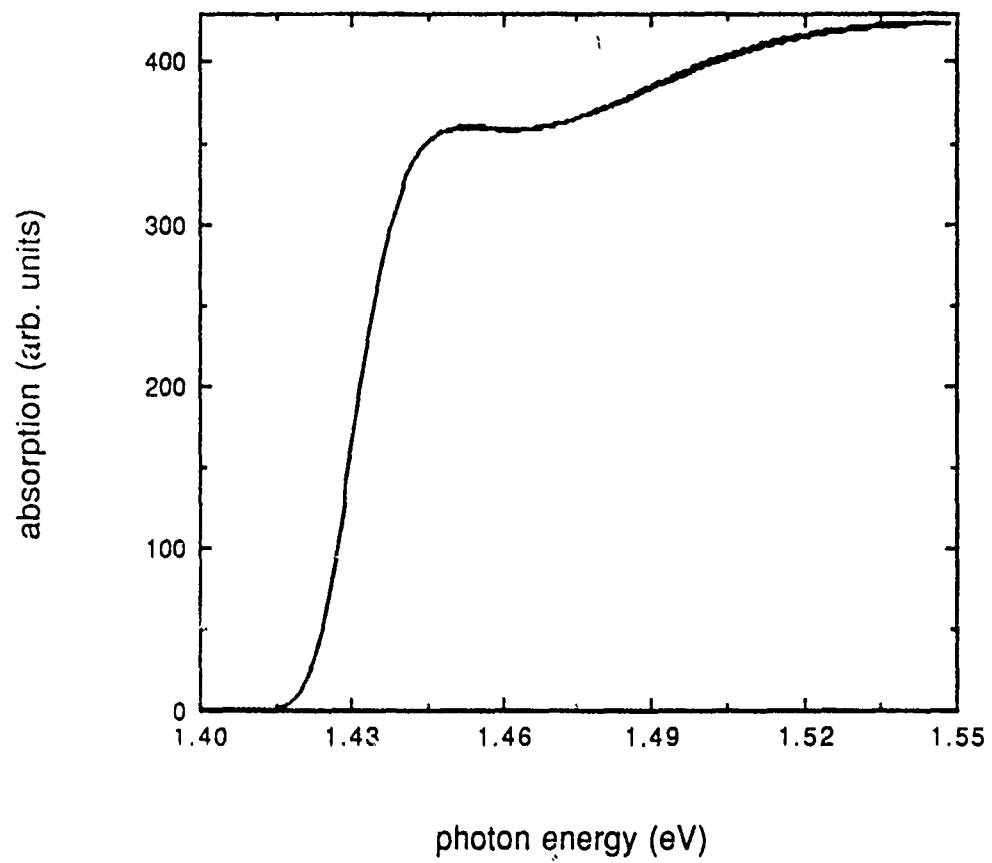


Figure 4

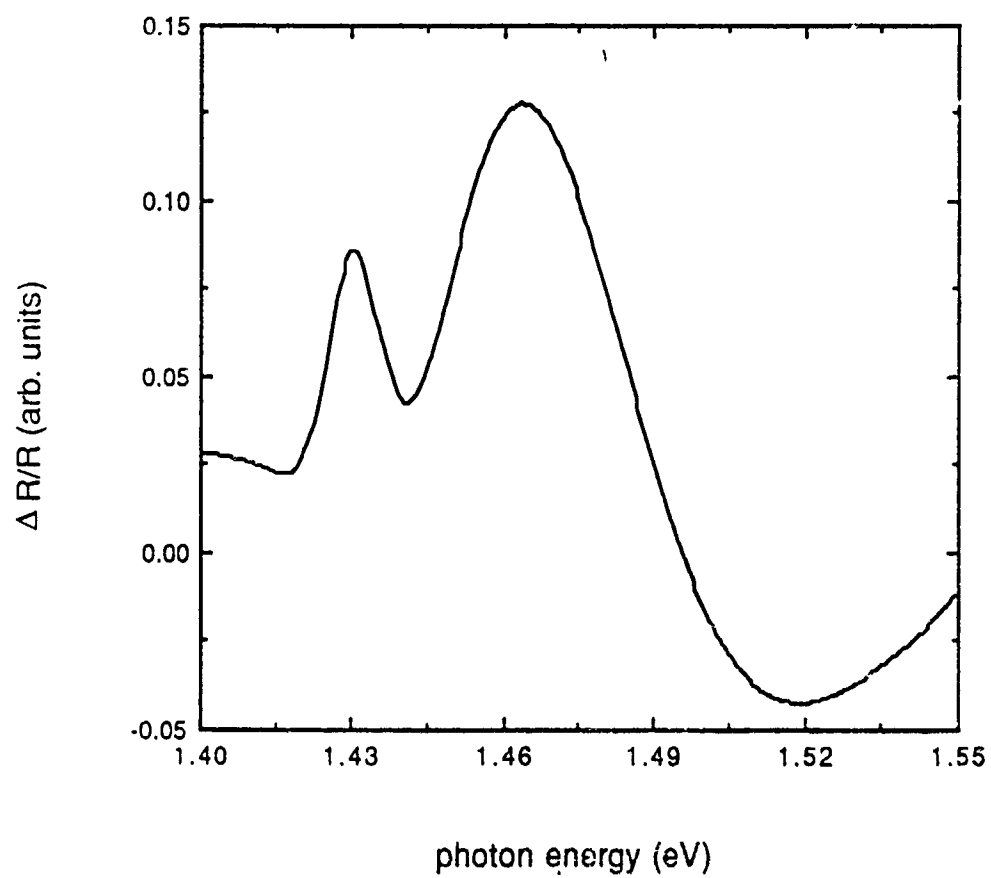


Figure 5

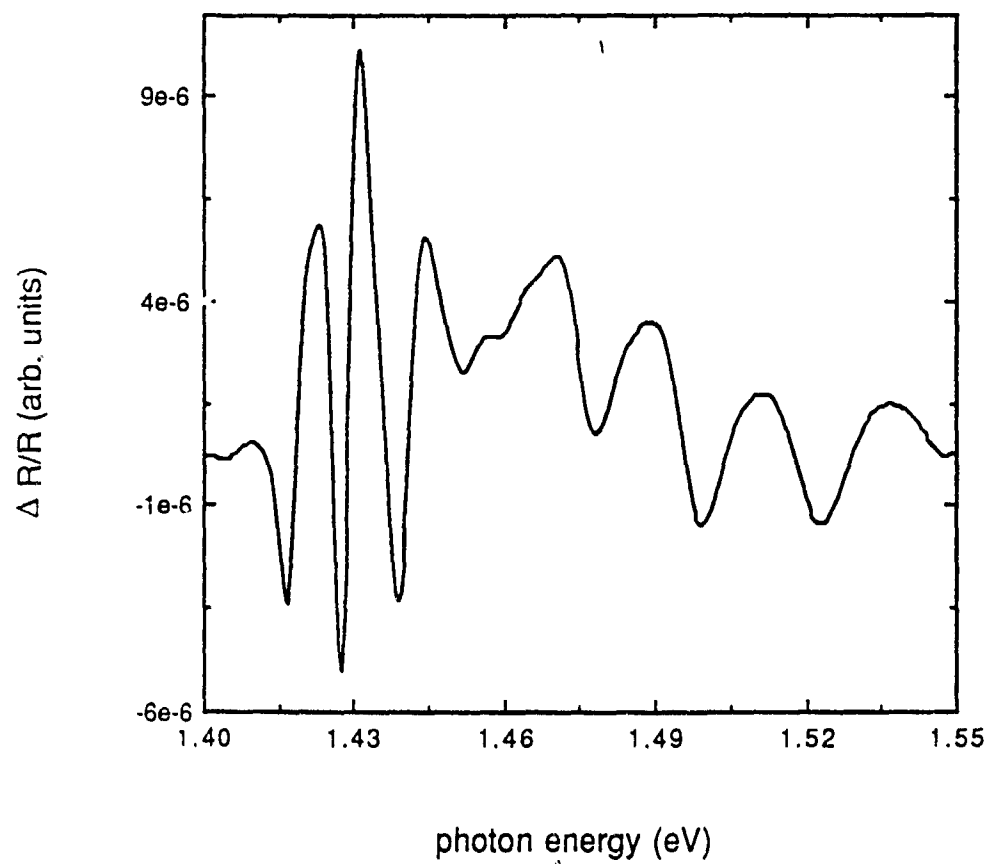


Figure 6

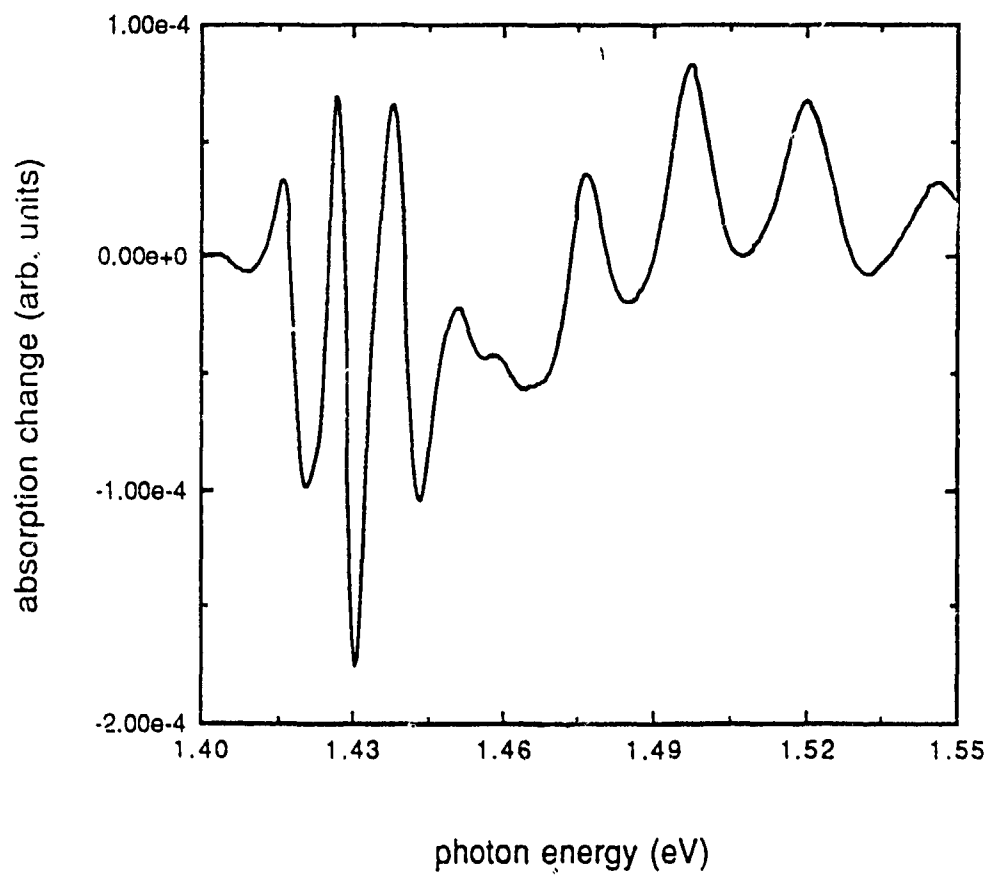
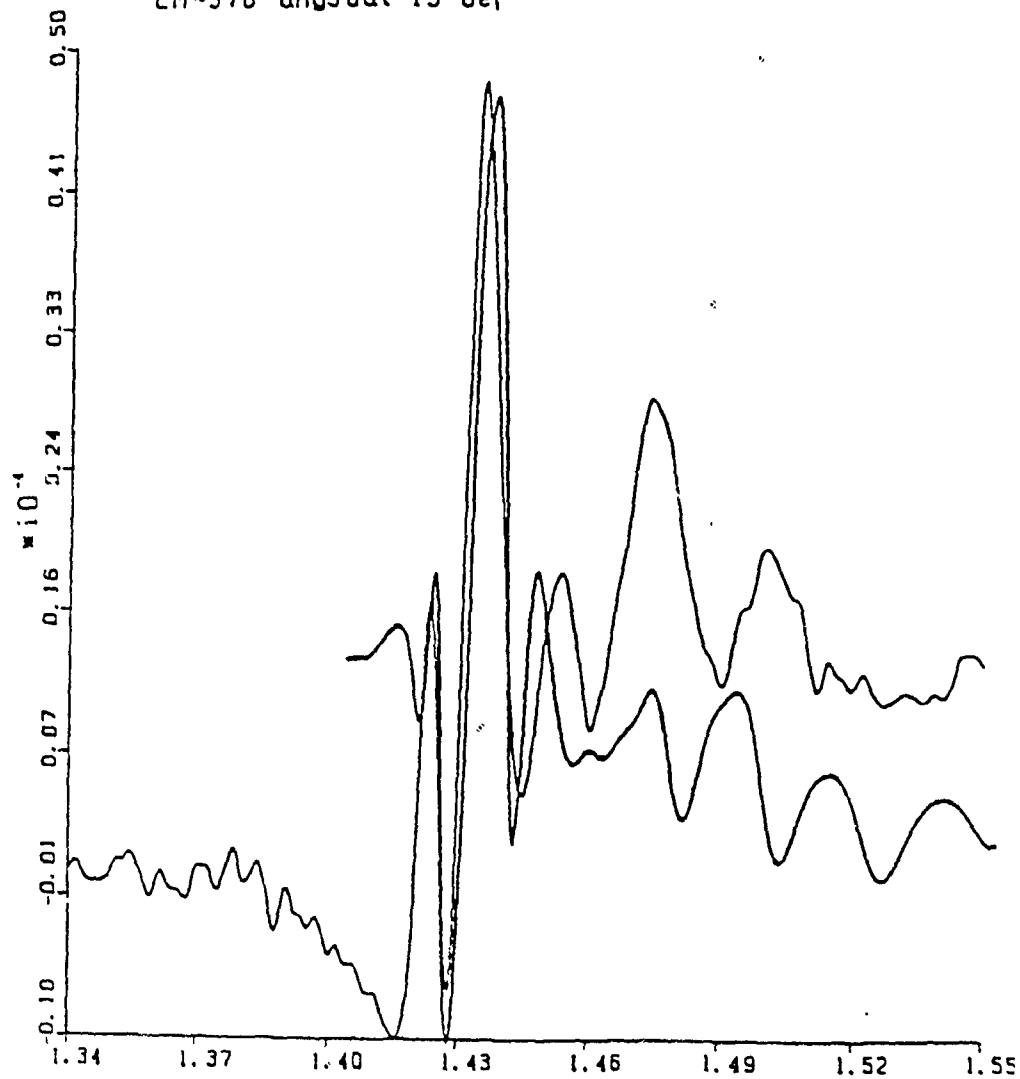


Figure 7

EM-376 any3dat 15 dei 7-5-91



PULSED LASER DEPOSITION AND NEAR OPTIMAL FORGING ACCOMPLISHMENTS

Pulsed Laser Deposition and
Near Optimal Forging Accomplishments
By: Douglas C. Moore

ABSTRACT

Accomplishments this Summer have been two fold. First, the pulsed laser deposition process presently being used by Wright Laboratories tribology group is being aided by the manufacturing research group in order to optimize the process. This optimization effort was advanced by software developments using the LabVIEW data acquisition and processing package from National Instruments. Also, preliminary investigation of possible quality improvements of the near-optimal-forging process were completed. Second, the near-optimal-forging program underway at Wright Laboratories seeks to use advanced computer modeling and insitu control parameter sensing to decrease forged-part variability while increasing product performance. The forging of high performance alloys has historically been difficult to control. With the use of insitu temperature measurements combined with advanced computer modeling, precise process paths can be generated and executed in order to produce consistently good product. This effort is being assisted by the manufacturing research group by investigating various methods of insitu process sensors. Particularly eddy-current and ultrasonic techniques have been researched for the purpose of noncontact temperature sensing of the preformed billet.

Introduction

Pulsed Laser Deposition Accomplishments:

The pulsed laser deposition (PLD) process under investigation at Wright Laboratories is being used to apply thin film coatings on a substrate. Such coatings are principally used as solid lubrication for applications where liquid lubricants would either be impractical or provide insufficient performance. There are many other applications for this process, and it is expected that industrial applications will use this technology on a production basis. Unfortunately, the process is difficult to sense and control since it is not well understood at present.

In order to fill this knowledge gap about the process dynamics, the manufacturing research group is presently acquiring and building a PLD chamber to be used specifically for process identification and control development. The existing chamber will be used by the tribology group for producing the films that they need for their research. Both chambers will use a common laser and control system developed around a MACINTOSH IIx computer. The main software package used to develop the control system is "LabVIEW," a product developed by National Instruments. This package is a high-level programming environment that does for scientists and engineers what the spread sheet program does for business people. It allows the researcher to quickly develop complex software for data acquisition and processing. Using programming languages such as "C" or "Pascal" can produce similar results but require many more hours of work to develop a usable application.

To lay the ground work, an overall structure for the LabVIEW-based application was developed. This structure is object-based which means that the application is divided into programming objects that contain the data and code needed to perform a specified task. This approach minimizes programming module interdependencies and generally makes the application easier to use and modify.

Near Optimal Forging Accomplishments:

Along with the PLD accomplishments, progress has been made in assisting the near optimal forging effort at Wright Laboratories. Some pioneering work is been done to integrating and automate the design and control of metal-forming processes. Advances in this area will lead to reduced lead times for new product design and delivery as well as improved product quality and performance.

To facilitate the application of this automated product/process design and control, various methods of insitu parameter sensing were investigated. One problem with the metal forming processes of interest is the difficulty with which they can be sensed. In order to achieve the most precise execution of a desired process path, accurate real-time information is required on the internal state of the process. Various methods of nondestructive testing were examined as possible techniques of data gathering that would help in reducing final product variability. These advanced sensing techniques offer the advantage of revealing information about the process formerly inaccessible. Eddy-current testing was arrived at as a practical method of insitu process measurement.

Pulsed Laser Deposition Discussion

In order to provide an application that is functionally sound, easy to modify, and user friendly an object-based structure must be employed. The LabVIEW environment presents some difficulties with regard to an object-based application. The main difficulty is that natural data dependencies that develop in any LabVIEW application. For example, if one process needs information from another, a wire is placed between them. This establishes a data dependency which forces the data supplying process to execute first before the data receiving process can execute. This simple arrangement forces the overall application to be strictly procedural. Multiple processes that share data must execute in sequence not simultaneously. This is incompatible with real-world applications that need to perform many tasks simultaneously and in a real-time fashion.

The solution is to create the LabVIEW application in an object-based style. By object-based I mean the following:

1. An object or "task" contains a globally accessible data storage structure and the code necessary to process it.
2. These tasks are written to consume a minimum amount of processor time. The processor must not spend much time servicing the task.
3. Complex tasks should function as state machines or sequencers.
4. The tasks will communicate via their global data structures similar to a network of independent devices.

A typical flow chart would be as follows:

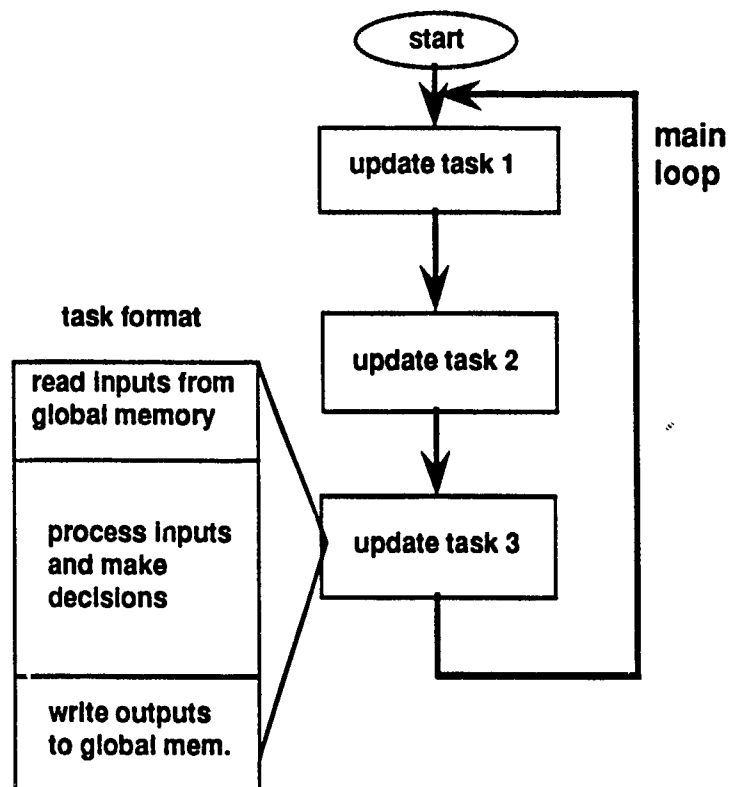


FIGURE 1. OBJECT-BASED FLOW CHART

Since each updating process requires only a short period of time to execute, the processor can service each task frequently so that the user sees all tasks executing in near real time. This results in CPU time being shared by the various tasks. Each time a task is serviced, it reads information from its own and other global storage structures as well as external and operator inputs. Once this information is obtained, the task can determine where it left off on the last service and what has changed in the environment. It can then make decisions about what outputs need to be changed and what the next state of the task should be.

An application may contain more than one main loop. For example, the PLD control project could be divided into separate modes of operation (such as: initialize, operator setup, run, and shutdown modes). Initialize would setup the system to operate and take care of certain housekeeping chores. Operator setup would be used to adjust the experimental apparatus properly before starting the experiment. Run would actually conduct the experiment, and shutdown would bring the system to a safe halt. The above object-based structure could be restricted to just the "run" mode where simultaneous tasks need to be executed. The tasks in the run mode would be as follows: mirror control, XTC (thickness gauge) control and data logging, vacuum monitor, laser control, data storage and analysis.

The benefits of this approach are as follows:

1. It minimizes interdependencies between the modules or VI's. This allows new tasks to be added without having to make major modifications to other tasks or to the structure of the application.
2. It makes the application event driven rather than data driven. This puts the operator in a more powerful position when using the application.
3. The application becomes more flexible and configurable.
4. It makes better use of CPU time by allowing tasks to execute simultaneously.
5. Tasks can also be prioritized so that low priority tasks do not delay important ones.

A hierarchy of a typical object-based application could be as follows:

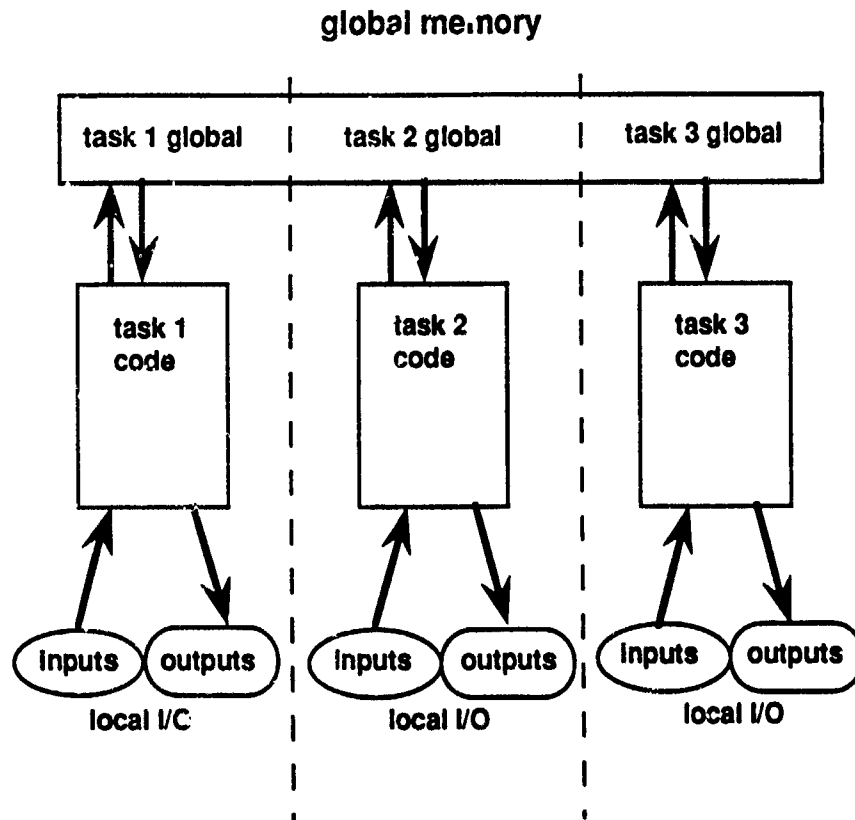


FIGURE 2. OBJECT-BASED HIERARCHY

The tasks have their own local I/O that they update at the time of servicing. Also, they communicate with other tasks and keep track of their own progress via global memory. Isolation of tasks is increased by allowing tasks to write only to their own reserved section of global memory. However, they can read any section of global memory. In this way data sharing takes place through a standardized conduit arrangement instead of haphazard connections between tasks.

Pulsed Laser Deposition Results

The PLD control program is divided into modes which represent the major operating states of the application. One can think of the application as a state machine where the modes are the various states. For example, when the application first starts, the operator will have to calibrate the software as to the position of the beam on the target. This gives rise to a "Jog Beam" mode where the operator will do the necessary calibration of the software with the experimental apparatus. After this initialization, raster patterns and other experimental methods could be chosen by the operator. This mode could be called "Pre-Experiment." The operator would use this mode to setup the application to execute the experiment in a certain way. The next mode might be the "Execution" mode. This state of the application executes the desired process path. After the experiment is executed, the gathered data can be further processed by the application's "Post Processing" mode. The last mode would be the "Shutdown" mode. This mode would bring the system to a smooth termination.

The method used by the application to go from mode to mode can be designed in a number of ways. A simple sequencer would probably be too restricting to the operator since it would force the modes to be executed in a fixed sequence. A more flexible method would be to incorporate next mode logic into each mode. This would be a set of rules used by the present mode to determine the next mode. For example, these rules might allow the operator to revert to the "Jog Beam" mode if the mirror apparatus gets out of calibration. It probably would not be advisable to allow the operator to go randomly from mode to mode, but a reasonable amount of flexibility would be useful. Each mode contains its own front panel which pops up when that particular mode is active. Only one mode can be active at any one time, and therefore, only one front panel is on screen at any one time. This keeps the number of front panel controls to a minimum which minimizes execution time for the

application. When the mode is terminated, its front panel pops down and the next mode's panel pops up. Various data structures called "queues" will be used to buffer communications between the tasks of any particular mode. For example, the Oriel laser mirror controller is a simple device that cannot accept more than one command at a time. If several commands need to be sent at a particular moment in the execution of an experiment they are stored on a "command queue" so that they can be passed on to the mirror controller in a organized fashion. This method can also be used to buffer commands from the operator. Suppose the operator enters several commands. These commands may not be immediately executable which would require that they be queued and executed in a logical fashion.

Near Optimal Forging Discussion

The present work being done to automate metal forming product/process design makes use of dynamic material models and finite element analysis to produce a desired process path plan. Once a process path plan has been determined from the dynamic material model and finite element analysis, accurate realization of this plan is largely limited by the equipment and controls in use. If theoretical reasoning is to be satisfactorily verified, the process must be properly under control. Advanced sensors is an area where control technology can aid the process of metal forging by providing up-to-date information on the process state. Typically, process control has taken the approach of controlling the process from its boundaries. For the typical forging process this might take the form of controlling the ram pressure, speed, and stroke. This approach is limited because the control system must infer the state of the process from the boundary conditions alone. For example, suppose variations exist between billets used in a forging process. Even if the

exact same boundary conditions were maintained for both, significantly different outcomes may result. Advanced sensing techniques can be used to view the inner workings of the process insitu. This will lessen the need to infer the state of the process and increase reliability and repeatability.

Advanced sensors is an area where control technology can aid the process of metal forging since work piece temperature is an important parameter in a forging process. The process stability maps for the aluminum alloys of interest show a sensitivity to temperature. In other words, knowing the temperature of die and work piece more accurately will help the control system maintain the quality and stability of the process. Unless we know the temperature of the billet accurately before forging, the probability of the process operating outside the desired stability window is greatly increased. Granted, one can infer the temperature from such data as furnace temperature, time in furnace, time left to cool at shop temperature, etc., but this method leaves a lot to be desired. In order to insure an accurate prediction of billet temperature, heating time for the billet must be lengthy, furnace temperature must be known exactly, and held very constant, so that no temperature gradients (or at least only highly predictable ones) exist in the billet.

A number of sensing alternatives were examined in order to find a suitable match for the processes of interest. Infra-red, acoustic, and eddy-current methods of thermal sensing were examined as techniques that could provide feedback on the process insitu. Infra-red methods seemed to be prone to inaccuracies and calibration problems, and it does not offer much information on the internal temperature of an opaque object. Acoustic and eddy-current methods, however, offer promise. Traditional temperature sensors like thermocouples would be impractical for measuring the temperature of the billet. However, eddy-current and ultrasonic techniques offer the possibility of yielding precise information about the temperature of the billet prior to forging without requiring permanent fixturing on the billet itself.

Eddy-current testing has been used by industry for years to improve quality control in both off and on-line systems, and it is considered to be the most usable at present for the forging processes of interest. It offers a number of advantages over other types of nondestructive testing like ultrasonic and X-ray testing in that the apparatus needed is usually less exotic. The added information from this technique will allow the forging process equipment to more closely produce the desired product by more closely following the temperature requirements of the process model.

Near Optimal Forging Results

Eddy-current testing has been used recently (Kahn, 1988) to measure the temperature of an aluminum extrusion line. The system actually measures the resistivity of the material just as it leaves the die. Since resistivity is directly related to temperature, the temperature was calculated and used as a closed-loop feedback signal to control the ram speed. Closed-loop temperature feedback was possible in this case since the process was a relatively slow one. Even though closed-loop temperature feedback is impractical for the forging process of interest because of time restrictions, a similar technique could be used to sense the temperature of a billet just prior to deformation. Eddy-current testing has also been used to obtain information on the precipitation hardening process for some aluminum alloys (Rosen, 1989) as well as heat treatment quality. Figure 3 shows a possible system. The primary coil is driven by a current regulated AC source, and the transfer impedance of the sensing coil is determined. This is directly related to the billet resistivity and consequently the billet temperature.

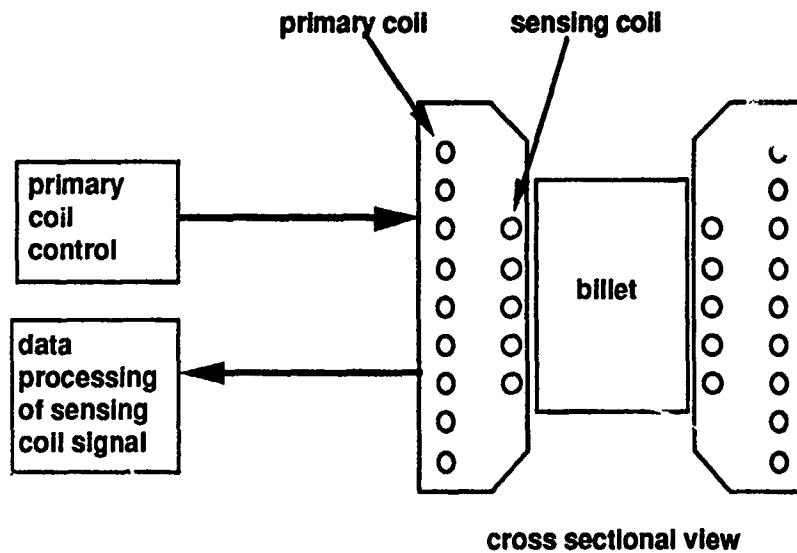


FIGURE 3. EDDY-CURRENT BILLET TEMPERATURE MEASUREMENT

Eddy-current testing can provide the needed temperature information in a noncontact way. The first step is to excite the conductive material under test by placing it within a alternating magnetic field. Eddy-currents are then induced in the material, and these currents are oriented orthogonally to the direction of the exciting magnetic field. They also have the same frequency as the frequency of excitation. These currents produce a magnetic field of their own which directly opposes the excitation field. The magnitude of these currents determine the strength of this counter magnetic field. The resulting composite field can be used to generate an electrical signal which upon decoding will reveal temperature information about the material under test since the material's temperature will directly affect the strength of the eddy currents. In the above mentioned system, the following technique was used:

1. Empirically determine the relationship between temperature and resistivity of the material .

2. Determine the resistivity of the test specimen by the following:
 - A. Determine angle α as defined below:
 1. Measure the complex transimpedance as defined below:
(Sensing coil voltage) / (Primary coil current)
 2. Do this for a band of frequencies ranging from 10Hz to 10KHz, and determine the knee of the resulting impedance curve.
 3. α is extracted as follows:

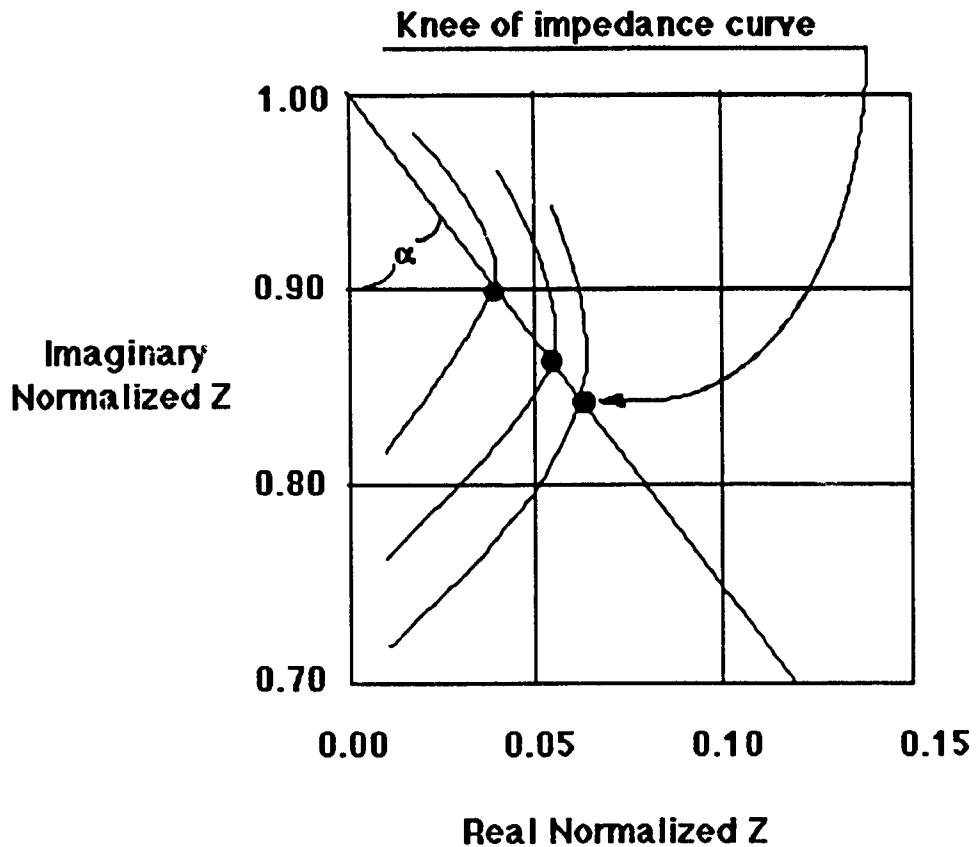


FIGURE 4. PLOT OF SENSOR COMPLEX IMPEDANCE

This shows three possible curves from different samples of the same material.

- B. α is used to determine $X = R\sqrt{\sigma\omega\mu_0}$. Uniform cylindrical specimens having similar values of α have the same value for X . The value X reveals the relationship between R (radius), σ (conductivity), ω (frequency), and μ_0 (permeability) for the specimen under test.
- C. Once the value X is known, resistivity can be extracted from the above equation at any frequency ω .
3. Finally, determine the temperature of the specimen from known empirically-determined relationship.

Kahn and Mester report a temperature measurement accuracy of ± 10 degrees Fahrenheit over a range of 950-1100 degrees F for aluminum alloys 6061 and 6063. They were also able to measure the diameter of a cylindrical specimen by extrapolating the impedance curve. The point where the extrapolated impedance curve crosses the imaginary Z axis indicates the fill factor (i.e., the amount of space inside the sensing coil filled by material).

Another method of obtaining temperature information on the billet would be the use of ultrasonic testing. H. N. G. Wadley (1986) and others have verified the use of ultrasonic time-of-flight measurements to obtain temperature profiles in heated metal samples. Ultrasonic emissions in the sample are produced by irradiating a spot on the sample surface with a pulse of high-energy laser light. This spot becomes a point source of ultrasonic waves which propagate through the sample. These waves are detected on the opposite side of the sample with an electromagnetic transducer which converts the acoustic wave into an electrical signal. One advantage of this technique over eddy-current methods is that it can provide information on temperature gradients in the sample. Moreover, eddy-current testing has as a depth limit. This depth limit varies with frequency and the material being sensed. The temperature profile pixels that can be generated by this ultrasonic method are represented as annular rings in figure 5.

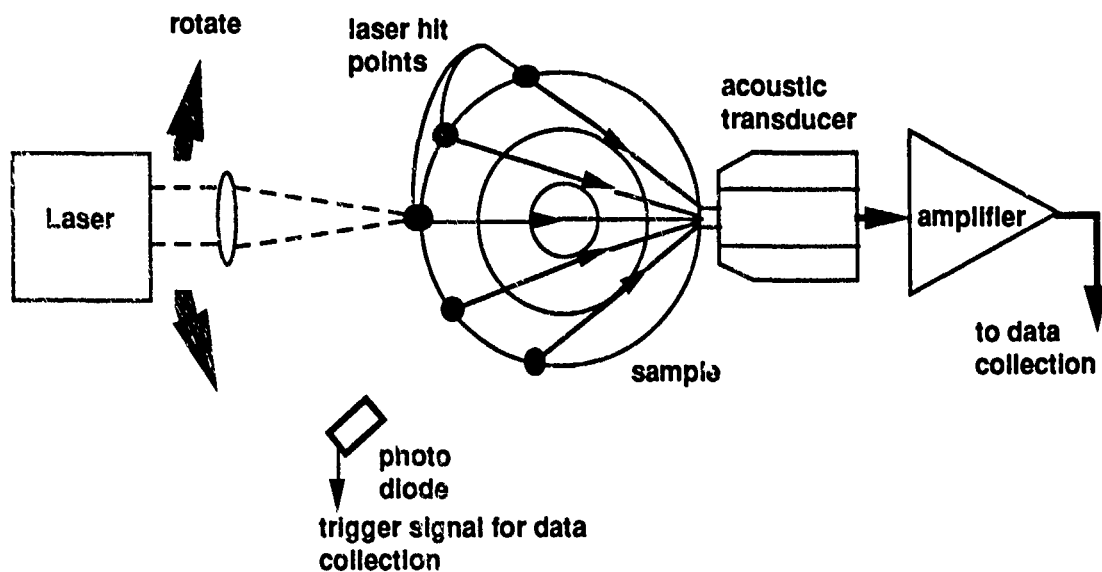


FIGURE 5. ULTRASONIC TIME-OF-FLIGHT TEMPERATURE MEASUREMENT

Conclusion

The two fold accomplishments in the pulsed laser deposition and near optimal forging processes have paved the way for future work. Now that the basic structure of the PLD control system has been defined, future software development can follow an organized path. Modifications and additions will be easier thanks to the object-based structure.

The ultrasonic method of temperature measurement is promising, but the simpler eddy-current method has been chosen as the most tractable of the two systems. It's simpler hardware will take less time to build and test, and provide an adequate temperature information for optimizing the forging process. Now that a sensing method has been chosen to improve the near-optimal-forging process (eddy-current temperature measurement), work can begin to actually implement such a system. Much research is needed regarding the hardware and software needed to make the eddy-current billet temperature measurement system actually work.

References

1. Kahn, Arnold H., Michael L. Mester. "An Eddy Current Sensor for the Measurement of Resistivity and Temperature of Aluminum Rod During Extrusion Processing". (Research Associate, The Aluminum Association, Inc.). Review of Progress in Quantitative Nondestructive Evaluation. Volume 7B. 1988. Pages 1599-1605.
2. Rosen, M. "Eddy Current Analysis of Precipitation Kinetics in Aluminum Alloys". Metallurgical Transactions A. Volume 20A. April 1989. Pages 605-610.
3. Wadley, H. N. G., S. J. Norton, F. Mauer and B. Droney. "Ultrasonic Measurement of Internal Temperature Distribution". Phil. Trans. Royal Society of London. A 320. 1986. Pages 341-361.
4. Wadley, H. N. G., et. al. "Eddy Current Measurement of Density During Hot Isostatic Pressing". (Research Associate, The Aluminum Association, Inc.). Review of Progress in Quantitative Nondestructive Evaluation. Volume 7B. 1988. Pages 1589-1598.

**Decomposition Pathways for Model Fluorinated Ether Lubricants
on Iron, Chromium and Nickel Surfaces**

Mary E. Napier
AFOSR Summer Research Fellow
Summer, 1991

Introduction

Perfluoropolyalkylethers (PFPE) possess several unique characteristics, thermal and chemical oxidative stability, wide liquid range, good temperature-viscosity character, chemical inertness and nonflammability, which make them attractive for use as high temperature lubricants.^{1,4} However, iron and titanium alloys, commonly found in the operating environments of the PFPE fluids, catalytically decompose these materials.^{2,8} With no metal present, the PFPE are stable up to 410°C in an inert environment and 316°C in an oxidizing environment.^{2,4,7} In the presence of metals, the thermal stability decreases dramatically in both environments.^{2,4,7} If these materials are to be used as proposed, it is extremely important to understand the nature of the catalytic degradation, so that structural modifications of the PFPE or surface modifications can be implemented to increase the working temperature range of the fluids, in the presence of metals.

Surface analytical techniques have been employed to study the PFPE liquid/metal interface and to investigate the interaction of PFPE with metals, under oxidation-corrosion and wear test conditions. For the oxidation-corrosion tests, the PFPE fluids were heated in the presence of metal coupons and the metal coupons examined using surface sensitive techniques.^{2,6,8,9} Decomposition of the PFPE fluids was characterized by the evolution of volatile, reactive degradation products, weight change in the metal coupon and the metal fluoride formation on the coupon surface. The initial step was postulated to be an interaction between the fluorocarbon ether and the metal, forming a metal fluoride, which catalytically degrades the PFPE.¹⁰ The catalytic activity of the metal fluoride was evidenced by the fact, that decomposition was immediately observed for PFPE in the presence of metal and metal fluoride, whereas in the presence of clean metal (no metal fluoride) an induction period was needed prior to the observation of decomposition.⁸ Under wear conditions, the reaction of PFPE was studied during sliding, in UHV (ultra high vacuum).¹⁰⁻¹² After the sliding concluded, the surface of the wear track was covered with reaction products; an adsorbed polymer layer over metal fluorides. The decomposition did not stop when the

sliding stopped. Reactive degradation products, proposed to be the metal fluorides, covered the wear track and the decomposition reaction continued for a short period of time, emphasizing the catalytic nature of the metal fluorides. The decomposition was not affected by the load and not the sliding speed. The formation of fresh, chemically reactive metal surfaces appeared to play the primary role in the decomposition, with the temperature secondary. The key step in the decomposition reaction, proposed from the oxidation-corrosion test and the wear test data, was the formation of metal fluoride which catalyzed the degradation.

The goal of this research is to expand on the body of knowledge available, on the PFPE interaction, with metals and to try answer more specific questions about the role of PFPE molecular structure and metal surface modification. Most of the work to date has been done investigating macroscopic phenomena using polymeric fluorinated ethers. Using these techniques, it can be difficult to ascertain the effect of subtle variations in molecular structural modification or surface modification. X-ray Photoelectron Spectroscopy was used to explore the surface chemistry of two model perfluoroalkylethers (PFAE) on clean and iron, chromium and nickel and oxidized iron polycrystalline surfaces, at temperatures between room temperature and 200°C. The model structures are monomer units of polymeric PFPE materials being investigated for use as high temperature lubricants. These lower molecular weight structures have the same functional units as the polymeric materials, can be easily handled and are simpler to interpret. The two model structures investigated were perfluorodioxalane, $\text{CF}_2\text{OCF}_2\text{CF}_2\text{O}-$, a ring structure and perfluorodimethoxy methane, $\text{CF}_3\text{OCF}_2\text{OCF}_3$, a linear structure. The three metals which were investigated are the main components of stainless steels, and comprise the bulk of the materials found in high temperature lubrication applications.

Experimental

The experiments were conducted using a Surface Science Instruments small spot XPS instrument equipped with a monochromatic x-ray source for sample analysis, a leak valve for introduction of gaseous samples and an Ar^+ ion sputter

gun for sample cleaning. The polycrystalline metal substrates were polished, then atomically cleaned in vacuum with cycles of sputtering and annealing. The model PFAEs were obtained from Exfluor. On site purification consisted of heart-cut distillation of the model PFAEs, followed by several freeze-pump-thaw cycles prior to introduction into the vacuum chamber. The oxidized iron surface was formed by exposure of a 300°C clean iron surface to 300 Langmuirs (1 L = 1×10^6 torr-sec) of air.

The experimental procedure involved exposure of a clean or oxidized metal surface, held at a constant specified temperature, to 1×10^7 L of one of the model structures. The temperature of the substrate was varied between room temperature and 200°C. XPS spectra of the substrates were acquired before and after exposure to the PFAE and changes in the substrate and the adsorbed PFAE layer were recorded, as a function of exposure and temperature.

Materials Characterizations

XPS is a surface sensitive, analytical technique capable of providing elemental analysis and oxidation state information. For example, typical carbon binding energies for the model fluorinated ethers are as follows:

CF ₃ O, 296.0 eV	CFO, 292.0 eV
CF ₂ O ₂ , 295.0 eV	CF, 289.0 eV
CF ₂ O, 294.1 eV	

Typical carbon binding energies for nonfluorinated ethers are 286.6 eV for CO and 285.0 eV for CH₄. Some system to system variation in the binding energies of the model PFAEs can be anticipated. As expected, the binding energies for the fluorinated compounds are substantially higher than for the nonfluorinated compounds. As the number of electronegative substituents increased, the binding energy increased and these carbons can be resolved using XPS. Two peaks are expected in the C(1s) spectra for intact perfluorodioxalane; CF₂O₂ at 295.0 eV and CF₂O at 294.1 eV in a 1:2 ratio. Two peaks are also expected in the C(1s) spectra for intact perfluorodimethoxy methane; CF₃O at 296.0 eV and CF₂O at 294.0 eV in a 2:1 ratio. In addition, the oxygen binding energies for the fluorinated ether is 535.5 eV compared to 531.5 eV for a nonfluorinated ether. The fluorine

binding energy for C-F an organic fluorine is 689.0 eV compared to 684.0 eV for an inorganic fluoride. Shifts in the binding energies for carbon, fluorine and oxygen to lower than expected values and variations in the expected ratios of the carbon peaks are characteristic of decomposition of the adsorbed layer.

Results

The first model structure investigated was the ring model compound, perfluorodioxalane, with polycrystalline iron, nickel and chromium substrates at room temperature and 200°C. Exposure of all three substrates, at room temperature and 200°C, to 1×10^7 L of perfluorodioxalane resulted in the deposition of 5 to 20 atomic percent of carbon and oxygen, observed at 285.0 eV and 532.0 eV, respectively and growth in 0 to 5 atomic percent of inorganic fluoride, observed at 684.5 eV. No correlation was observed when comparing the amount of carbon and oxygen deposited with the exposure nor with the carbon oxygen ratio of the ring model structure. In addition, no correlation was observed in the amount of carbon oxygen deposited from run to run. The carbon oxygen layer was severely deficient in fluorine.

The second model structure investigated was the linear model compound, perfluorodimethoxy methane with polycrystalline clean iron, at room temperature, 80°C and 200°C and oxidized iron at room temperature, 80°C, 150°C and 200°C. Figure 1 shows the C(1s) spectra for the mPFAE exposed a) clean iron and b) oxidized iron, at room temperature. A progression of carbon binding energies is observed, corresponding to the following species; CF (289.0 eV), CO (286.1 eV), carbidic carbon (284.8 eV) and graphitic carbon (283.4 eV). As expected, no graphitic carbon is observed on the oxidized iron surface. At similar exposures, less carbon is deposited on the oxidized iron surface than on the clean iron surface. Figure 2 shows the F(1s) spectra for the mPFAE exposed a) clean iron at room temperature, 80° and 200°C and b) oxidized iron at room temperature, 80°C, 150°C and 200°C. The fluorine species observed at 684.5 eV corresponds to FeF_x , an inorganic fluoride. Concomitant with the carbon deposition, metal fluoride growth was less on the oxidized surface than on the clean surface. Additionally, as the temperature of the substrate was increased, the carbon deposition and the

metal fluoride growth decreased, but again differences were seen when comparing the clean and the oxidized iron surfaces. On the clean iron surface, carbon deposition and metal fluoride growth decreased dramatically as a function of temperature, with no carbon nor metal fluoride observed, following the exposure of the 200°C surface. Whereas on the oxidized surface, the decrease was gradual, with carbon and metal fluoride observed at all temperatures up to and including 200°C.

Discussion

The lack of correlation for the ring structure indicated that the surface chemistry observed was unrelated to the major component of the gas introduced into the chamber and was related to the impurities in the gas. The perfluorodioxalane did not react with the polycrystalline Ni, Cr and Fe surfaces at room temperature and 200°C.

The XPS spectra acquired for the perfluorodimethoxy methane exposed iron surfaces, is not the spectra anticipated for the adsorption of intact linear mPFAE. For both the clean and the oxidized iron surfaces, the progression of low XPS binding energies observed in the carbon spectra and the formation of metal fluoride indicated that decomposition occurred and appeared to proceed via a defluorination of the carbon oxygen backbone. This decomposition pathway is similar to the pathway of hydrocarbon ethers on transition metal surfaces, where the rate limiting step is the scission of the C-H bond.^{13,14} For the case of mPFAE decomposition, the rate limiting step is the scission of the C-F bond.

A comparison of the bonding of hydrogenated and fluorinated ethers on Ru(001) and Ru(100) reveals that hydrogenated ethers decompose on clean Ru surfaces, as expected, but fluorinated ethers do not decompose,^{15,16} contrary to the present results on iron. The difference between Fe and Ru can be understood in terms of bond energy considerations. The bond energy for Ru-F, calculated from $\Delta H_f(\text{RuF}_3)$, is 256 kJ/mol, significantly lower than either the Fe-F bond energy, 480 kJ/mol, or the C-F bond energy, 491 kJ/mol.^{17,18} PFAE decomposition is approximately thermoneutral on iron and endothermic on Ru. This comparison also implies that the formation of iron fluoride is the thermodynamic driving

force for C-F bond scission.

Oxidative breakdown has been observed for polymeric PFAE fluids in the presence of oxidized metals.¹⁹ The proposed reaction theorizes that the acidic surface is an effective receptor for available electron density from the PFAE chain and the reaction proceeds via an interaction between acidic sites on the oxidized surface and the basic sites on the molecular chain. The decomposition of the mPFAE on the oxidized surface, appears to proceed through the same pathway as on the clean surface, differs in two respects. First the metal fluoride growth and the carbon deposition is decreased when compared to the clean. This implies that metal sites are controlling the chemistry of the mPFAE decomposition, even on the oxidized surface. On the oxidized surface fewer metal sites are available, so less decomposition is observed. Secondly, the decomposed mPFAE binds more strongly to the oxidized surface, with carbon deposition and metal fluoride growth observed at higher substrate temperatures. The enhanced Lewis acidity of the oxidized surface²⁰ and the increased chemisorption bond strength of the decomposition products implies that the mPFAE possesses a Lewis base character.

The linear mPFAE, perfluorodimethoxy methane, decomposed on the clean and oxidized iron surfaces, while the ring mPFAE, perfluorodioxalane, did not decompose on the clean iron surface. The decomposition of the linear model structure has been postulated to proceed through attack by iron on the C-F bonds of the terminal methoxy group.²¹ One possible explanation for the lack of decomposition of the ring structure is that no end groups are available for attack by the iron surface. The ring structure sterically protects the C-F bond from attack by the iron surface.

CONCLUSIONS

The linear structure, perfluorodimethoxy methane decomposes on both oxidized and clean iron surfaces, while the ring structure, perfluorodioxalane did not decompose. The decomposition proceeded through defluorination of the carbon oxygen backbone, with deposition of carbon, oxygen and growth in metal fluoride.

REFERENCES

1. Gumprecht, W.H., ASLE Trans. 9, 24 (1966).
2. Sianesi, D., Zamboni, V., Fontanelli, R. and Binaghi, M., Wear 18, 85 (1971).
3. Snyder, C.E. and Dolle, R.E., Jr., ASLE Trans. 19, 171 (1976).
4. Snyder, C.E., Gschwender, L.J. and Tamborski, C., Lubrication Eng. 37, 344 (1981).
5. Paciorek, K.J.L., Kratzer, R.H., Kaufman, J. and Nakahara, J.H., J. Appl. Poly. Sci. 24, 1397 (1979).
6. Jones, W.R., Jr., Paciorek, K.J.L., Ito, T.I. and Kratzer, R.H., Ind. Eng. Chem. Prod. Res. Dev. 22, 166 (1983).
7. Jones, W.R., Jr., Paciorek, K.J.L., Harris, D.H., Smythe, M.E., Nakahara, J.H. and Kratzer, R.H., Ind. Eng. Chem. Prod. Res. Dev. 24, 417 (1985).
8. Chandler, W.L., Lloyd, L.B., Farrow, M.M., Burnham, R.K. and Eyring, E.M., Corrosion 36, 152 (1980).
9. Carre, D.J. and Markowitz, J.A., ASLE Trans. 28, 40 (1985).
10. Carre, D.J., ASLE Trans. 29, 121 (1986).
11. Mori, S. and Morales, W., Wear 132, 111 (1989).
12. Mori, S. and Morales, W., Trib. Trans. 33, 325 (1990).
13. Rendulic, K.D. and Sexton, B.A., J. of Catalysis 78, 126 (1982).
14. Sexton, B.A. and Hughes, A.E., Surf. Sci. 140, 227 (1984).
15. Walczak, M.M. and Thiel, P.A., Surf. Sci. 224, 425 (1989).
16. Walczak, M.M., Leavitt, P.K. and Thiel, P.A., Trib. Trans. 33, 557 (1990).
17. Eneleus, H.J. and Sharpe, A.G., "Advances in Inorganic Chemistry and Radiochemistry"; Vol 24, Academic Press: New York, 1981.
18. Porter, H.A., Greenberg, E. and Hubbard, W.N., J. Phys. Chem., 69, 2308 (1965).
19. Zehe, M.J. and Faut, O.D., Trib. Trans., 33, 634 (1990).
20. Stair, P.C., J. Am. Chem. Soc., 104, 4044 (1982).
21. Napier, M.E. and Stair, P.C., submitted J. Am. Chem. Soc.

Figure Captions

Figure 1: C(1s) spectra for perfluorodimethoxy methane exposed room temperature a) clean iron and b) oxidized iron surface.

Figure 2: F(1s) spectra for perfluorodimethoxy methane exposed a) clean iron surface at room temperature, 80°C and 200°C and b) oxidized iron surface at room temperature, 80°C, 150°C and 200°C.

Figure 1

a) clean iron

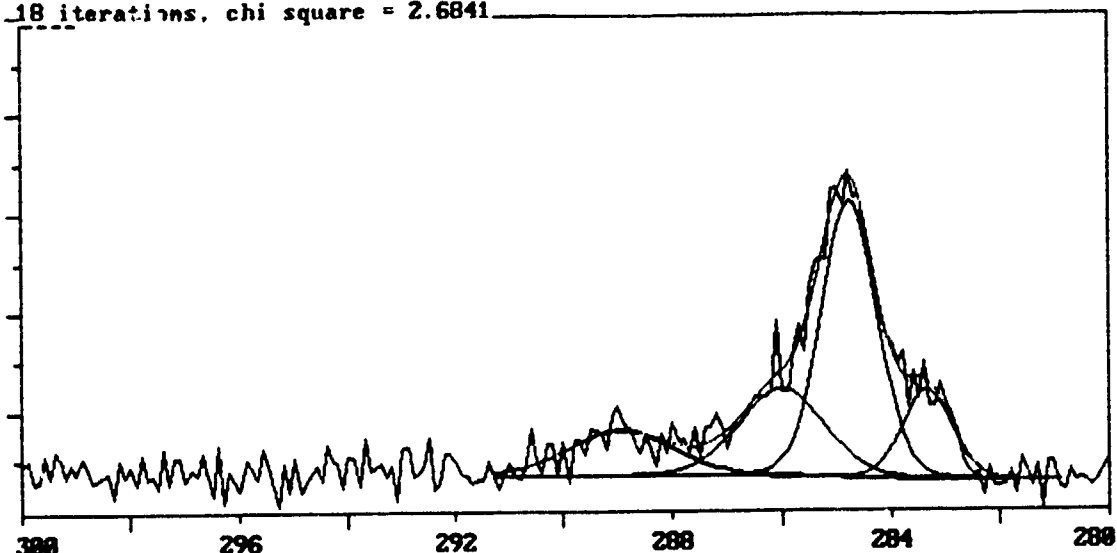
M-Probe ESCA Console User ID: IRON

Filename	Spot	Res	Flood eV	Scans	Description
FE_AFT8.MRS	288x750μ	2		20	clean poly fe. exposed to 1E7 L of C

Baseline: 291.38 to 288.98 eV

#	Position (eV)	Width (eV)	Area (cts)	Percentage (%)
1	286.88	1.81	3585.75	23.17%
2	289.88	2.23	2229.15	14.73%
3	284.83	1.19	7277.22	48.89%
4	283.41	1.87	2119.67	14.81%

18 iterations, chi square = 2.6841



b) oxidized iron

M-Probe ESCA Console User ID: IRON

Filename	Spot	Res	Flood eV	Scans	Description
FE_OXRT2.MRS	288x750μ	2		20	oxidized iron. room temp. exposed to

Baseline: 292.88 to 281.58 eV

#	Position (eV)	Width (eV)	Area (cts)	Percentage (%)
1	284.73	1.52	2858.28	68.57%
2	287.85	1.42	1884.36	23.84%
3	289.87	1.69	778.86	16.38%

12 iterations, chi square = 4.8588

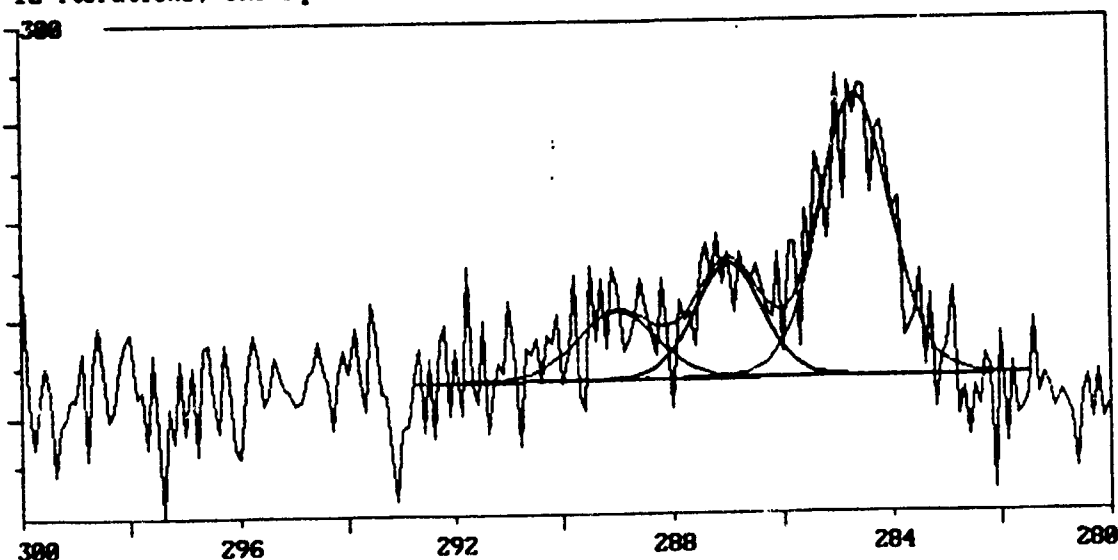


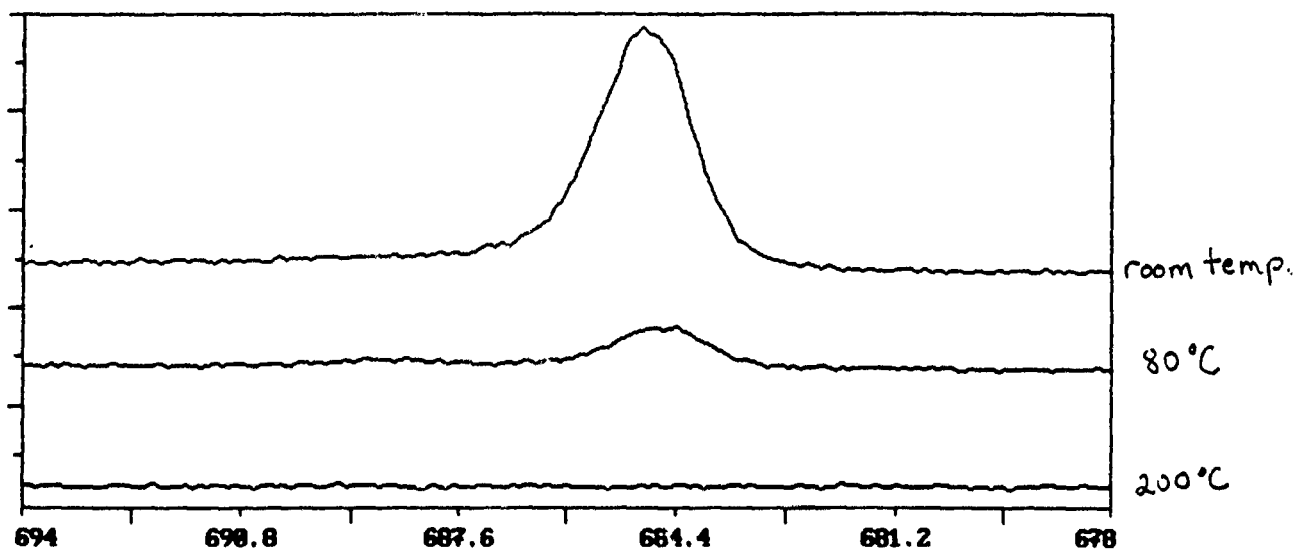
Figure 2

a) clean iron

M-Probe ESCA Console

User ID: IRON

Filename	Spot	Res	Flood eU	Scans	Description
FE_AFT11.MRS	200x750μ	2		20	fe. heated to 200 C. exposed to line
FE_AFT12.MRS	200x750μ	2		20	fe. heated to 80 C. exposed to line
FE_AFT8.MRS	200x750μ	2		20	clean poly fe. exposed to 1E7 L of C

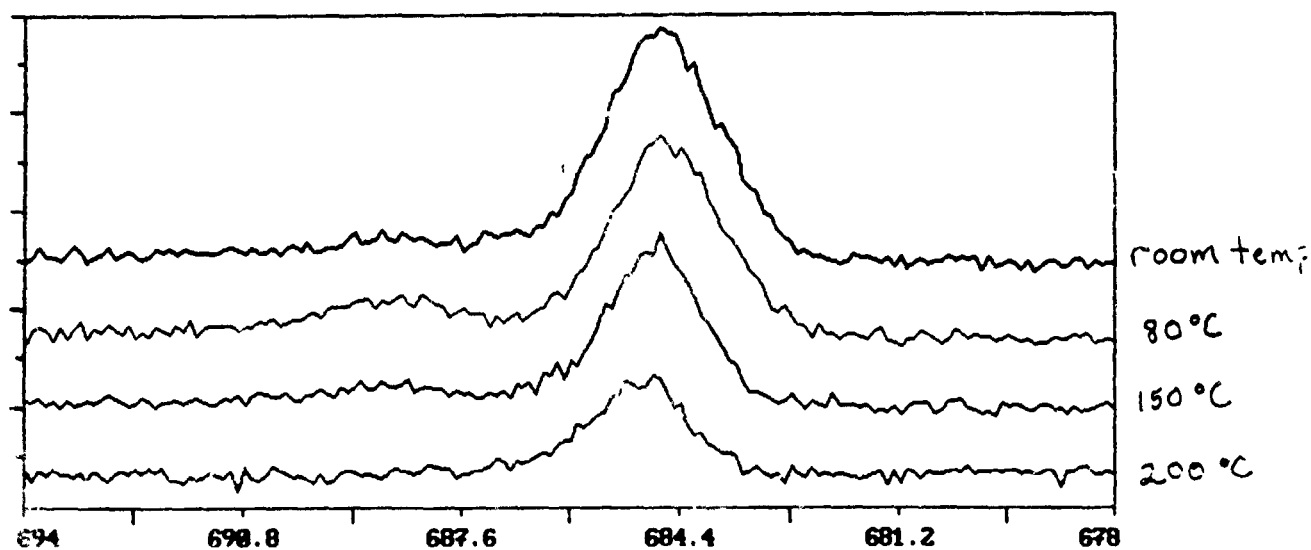


b) oxidized iron

Probe ESCA Console

User ID: IRON

Filename	Spot	Res	Flood eU	Scans	Description
FE_0043.MRS	200x750μ	2		20	oxidized iron, exposed to linear PFA
FE_0042.MRS	200x750μ	2		20	oxidized iron, exposed to linear PFA
FE_01E7H.MRS	200x750μ	2		20	oxidized iron, exposed to PFAE, line
FE_00RT2.MRS	200x750μ	2		20	oxidized iron, room temp, exposed to



ULTRASONIC BEAM PROPAGATION: DIFFRACTIONLESS BEAMS AND
BEAMS IN ANISOTROPIC MEDIA

Byron Newberry, Assistant Professor
Mark Preischel, Graduate Student
University of Cincinnati

ABSTRACT

Two problems were addressed during the summer research activity, both dealing with the propagation of ultrasonic beams as used in the nondestructive evaluation of materials. First, the possibility of making a transducer which produces an ultrasonic beam that does not diffract, or spread, as it propagates was explored. As a result of the work, such a beam was produced and a proposal for the construction of a practical device was developed. The second topic concerned the modeling of ultrasonic transducer beam propagation through anisotropic materials, such as fiber reinforced composites. As a result, simple formulas which predict the on-axis field profiles of beams in such materials were derived and validated through comparison with more complex models.

This report is broken into two sections describing each of these topics in detail.

I. PRODUCTION OF A DIFFRACTIONLESS ULTRASONIC BEAM

INTRODUCTION

The propagation of ultrasonic beams is a phenomenon of widespread interest to a variety of technologies including sonar, medical ultrasound, and nondestructive evaluation. One goal in most applications is the production of a narrow, highly collimated beam of sound. Rigid piston radiators have often been employed and have been thoroughly analyzed. This type of source has the generally undesirable attributes of a complicated near field interference structure as well as far field side lobes. Sources which produce a Gaussian amplitude distribution have been studied since, for this case, the previous disadvantages are eliminated. Unfortunately, Gaussian radiators are more difficult to manufacture [1,2]. Various types of focusing probes have also been analyzed for concentrating the sound in a narrow band over a short depth of field. Conically focussed, or axicon, probes have been examined for the purpose of extending the focal region for resolution over a greater depth of field.

One disadvantage common to all of the above sources, and indeed to any physically realizable source, is the phenomenon of beam spread due to diffraction.

Recently, in the optics literature, Durnin [3,4] has pointed out the existence of an axially symmetric harmonic solution to the scalar wave equation in the form of a transverse Bessel function modulated by an axially varying phase, $\phi(r,z) = J_0(\alpha r) \exp(j\beta z)$, where $k^2 = \alpha^2 + \beta^2$, and k is the wave number. When $\alpha=0$, this is just a plane wave solution. However, for $0 < \alpha < k$, we have a forward propagating field which has a beam-like transverse profile that does not vary with axial position, and is therefore diffractionless. The beam energy is not localized, however, but is instead infinite due to J_0 not being square integrable. Durnin analyzed a finite aperture approximation to a J_0 beam for the optical case and has shown both analytically and experimentally that a beam may be produced which can have a small spot size while retaining a significant depth of field over which the transverse profile is approximately constant. Initial attempts to produce a diffractionless acoustic beam have been reported by other investigators [5-7]. These have employed transducer arrays or transducer disks with radially varying poling or excitation. In this paper, we discuss an approach for the acoustic case analogous to the

method of Durnin and report the experimental generation of a nondiffracting J_0 Bessel beam. We also will discuss the potential for constructing a practical transducer via this approach.

EXPERIMENT

If the Bessel function solution is decomposed into an angular spectrum of plane waves, we find that the only wave vectors which contribute lie on the surface of a cone centered about the z axis and making an angle $\theta = \arcsin(\alpha/k)$ with it. Durnin showed, for the optical case, that this plane wave spectrum can be generated by a construction similar to that shown in Fig. 1, which is for the current acoustical case. A circular PZT piston transducer of radius a is used as the source. The center of the transducer disk is blocked by a circular disk of radius $b < a$ made of a highly attenuative material (in this case, polystyrene). The resulting effective source is a thin annulus.

An acoustically convergent lens of focal length F and diameter $2R$ is placed a distance F in front of the transducer. Each point on the ring radiates a spherical wave which the lens transforms into a plane wave traveling at angle $\theta = \arctan(a/F)$ relative to the z axis. Summing contributions from each point on the ring yields the

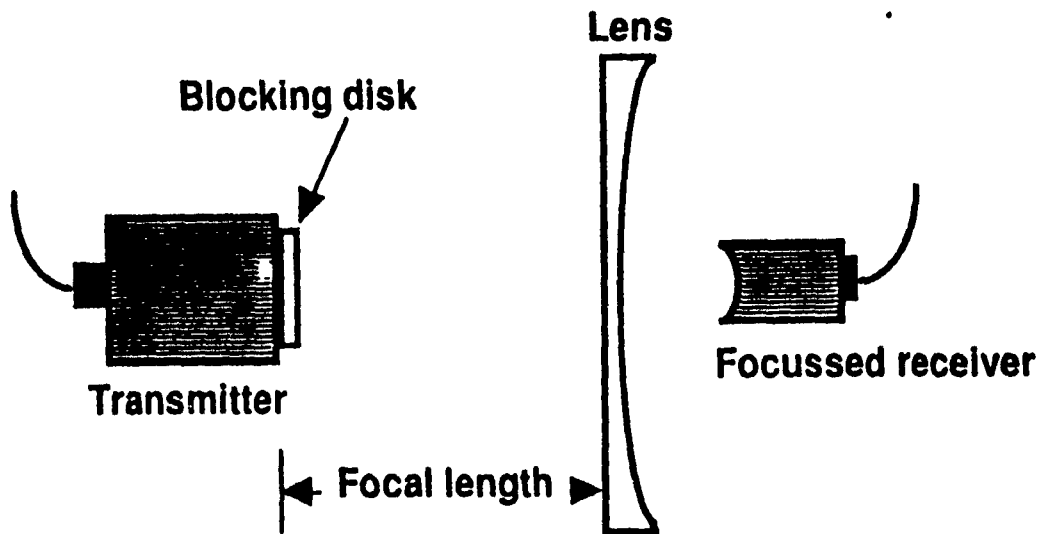


Figure 1. Experimental setup for producing diffractionless ultrasonic beam.

conical set of wave vectors necessary for a Bessel beam with width parameter $\alpha = (ka/F)(1+a^2/F^2)^{-1/2}$. If $a \ll F$, then $\alpha = ka/F$. The maximum possible depth of field is $z_{\max} = R/\tan\theta = RF/a$. Past this point, a shadow zone is produced due to the finite lens aperture.

A 0.25 inch radius, broadband transducer was used as the source. The blocking disk had a 0.22 inch radius. The lens was a 1.0 inch radius, fused quartz plano-concave which has an acoustical focal length of 6.85 inch in

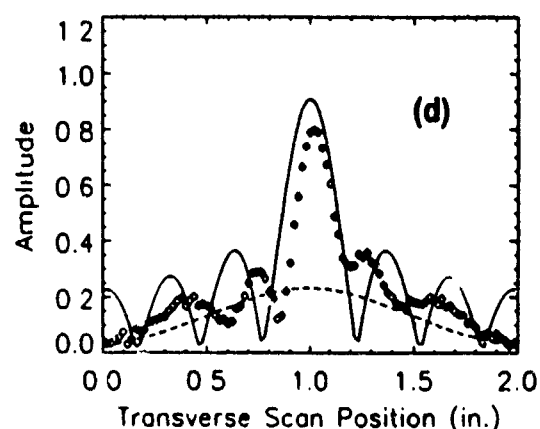
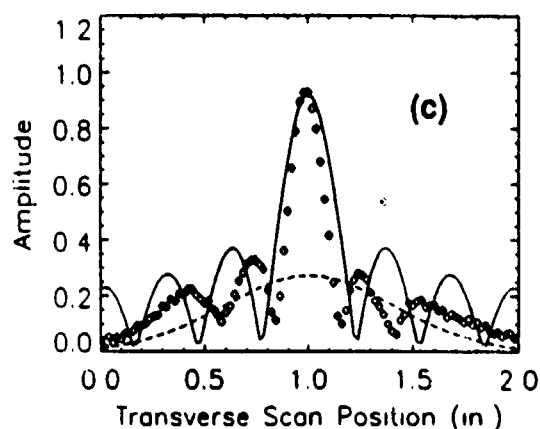
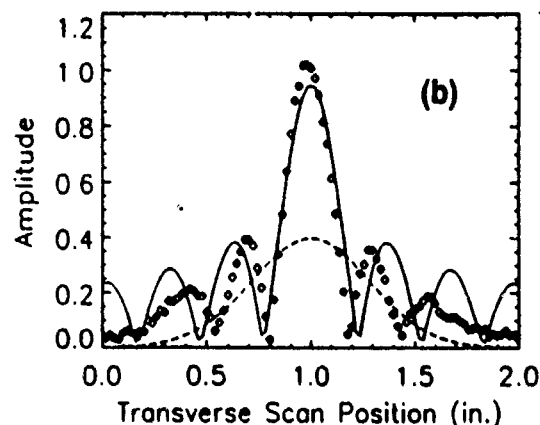
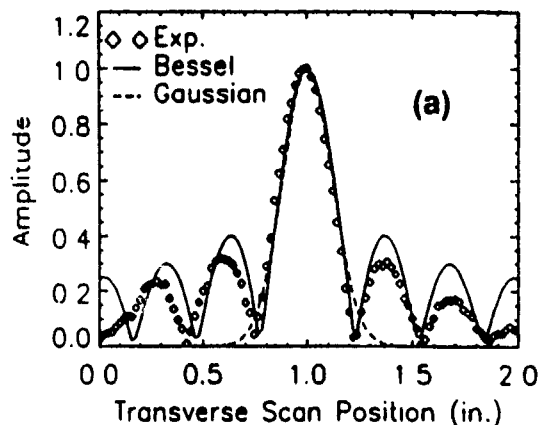
water. For measurement of the generated sound field, a 0.75 inch focal length, 0.25 inch radius transducer was scanned in a 2-D grid perpendicular to the beam axis. The focal spot of the receiver was much smaller than the Bessel beam width so that the measured amplitude profile should closely represent the actual Bessel beam profile at the axial distance corresponding to the focal point of the receiver.

The received rf waveform at each point in the scan grid was recorded and Fourier transformed to obtain the frequency spectrum. The amplitudes at particular frequency values were extracted for each point and plotted versus position in the scan grid, resulting in images of the transverse beam profiles at different frequencies and axial positions.

Figure 2a shows the data along a slice through the peak of the sound field transverse to the beam axis at 3 MHz, at an axial distance of $z=0.5$ inch past the lens. Also plotted is a J_0 function with $\alpha=10.4/\text{inch}$. This value of α , which gives a best fit to the data, corresponds to a ring radius of 0.224 inch, approximately the inner radius of the annulus. For comparison, a Gaussian beam with the same $1/e$ width as the main lobe of the Bessel function is also plotted ($w=0.17$ inch).

Figures 2b-d show the beam profile at distances of $z=10, 15,$ and 17.5 inches past the lens. The theoretical Bessel function has been scaled down at each axial distance to account for sound attenuation in water. It is apparent that the Bessel beam retains its initial width and amplitude to about 17.5 inches, although the side lobe structure has decayed. At 20 inches, which is not shown, the main lobe has broken down. The ideal z_{\max} for this case is 27.4 inches. At 17.5 inches, the Gaussian beam has decreased by approximately 80% from its initial amplitude. The Rayleigh range for the Gaussian is $z_r = \pi w^2 / \lambda = 4.6$ inch. The Bessel beam has propagated more than 3 times this distance without significant diffraction loss.

Because the source is a finite width annulus rather than an infinitesimal ring, a range of source radii contribute to the beam field, each producing a slightly different beam width parameter α . Since the phase parameter β varies with α , the wave becomes dispersive when the source has a finite width. This contributes to the breakdown of the Bessel profile before the ideal z_{\max} is reached.



2. Transverse profiles of Bessel beam at 3 MHz at axial distances of (a) 0.5 in., (b) 10 in., (c) 15 in., (d) 17.5 in.

The above results are quite exciting, although one disadvantage of the Bessel beam for the monochromatic case is the large magnitude of the side lobes. The beam is found to also be nondiffracting for the pulsed case. Figures 3a-d show the 2-D transverse profile scans of the peak-to-peak voltage of the rf waveform at various

distances from the lens. The pulsed beam is actually somewhat narrower than the 3 MHz case due to the contribution of higher frequencies. Also, the pulsed beam displays no sidelobes due to incoherent addition of the sidelobes at the different frequencies. Since ultrasonic scanning is often done in the pulsed mode, this shows promise for having a narrow, sharply peaked interrogating beam with a large depth of field.

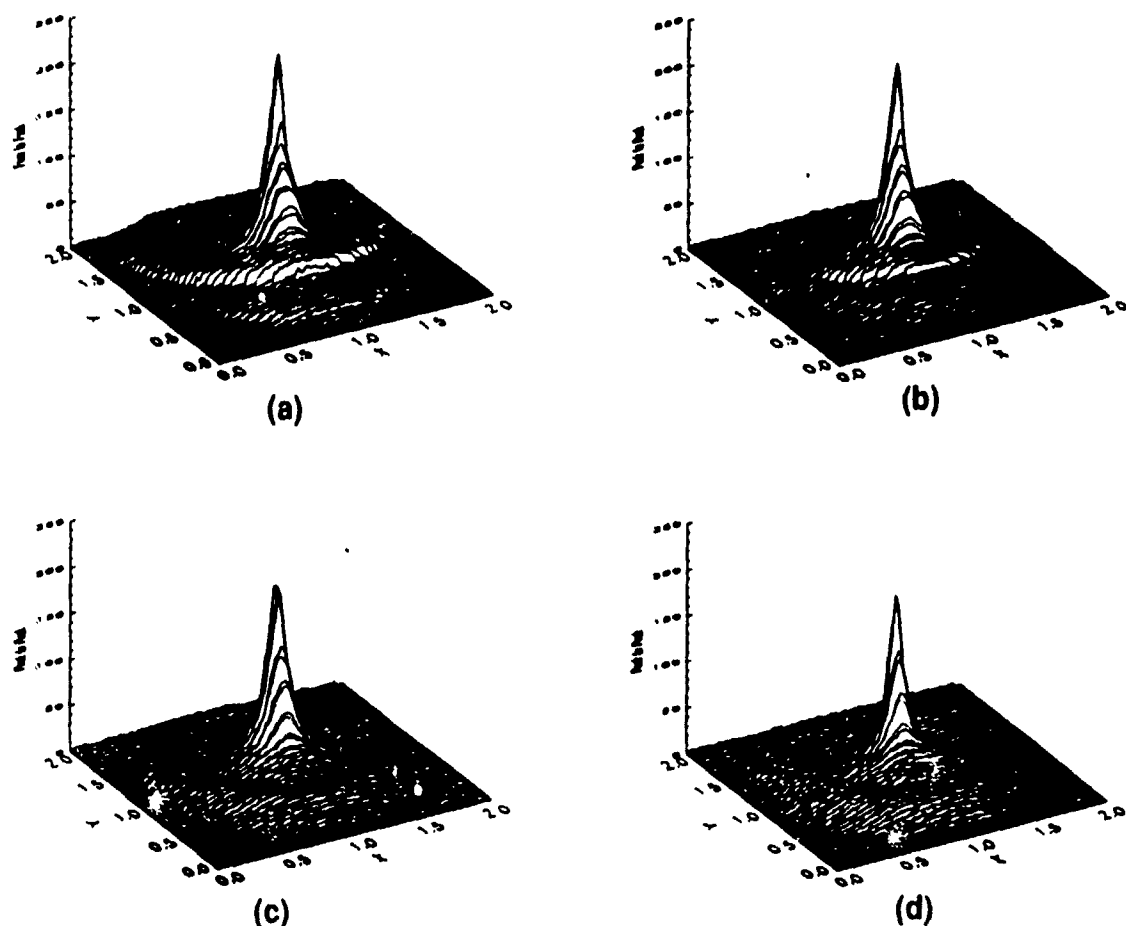


Figure 3. Transverse profiles of pulsed peak-to-peak voltage
(a) 0.5 in., (b) 5 in., (c) 10 in., (d) 15 in.

DISCUSSION

These results suggest that current ultrasonic transducer beam types can be greatly improved upon by the implementation of the diffractionless Bessel beam idea. The apparatus illustrated in Fig. 1, however, was contrived for the purpose of demonstrating the concept and is clearly not a practical one for ultrasonic inspection. However, it may be possible to construct a Bessel beam transducer by appropriately constructing a transducer element and lens combination. The manufacture should not be very different from that of a conventional focussed probe.

A proposed configuration is shown in Fig. 4. In this case, an annular transducer element is mounted on the planar surface of a thick plano-concave lens. The parameters such as annular radius and width, lens thickness, lens diameter, and lens radius of curvature must be chosen using thick lens theory to ensure that the annulus rests at the focal depth of the lens and that the desired characteristics of the beam are achieved. Further work on this subject will include the construction and testing of this transducer configuration.

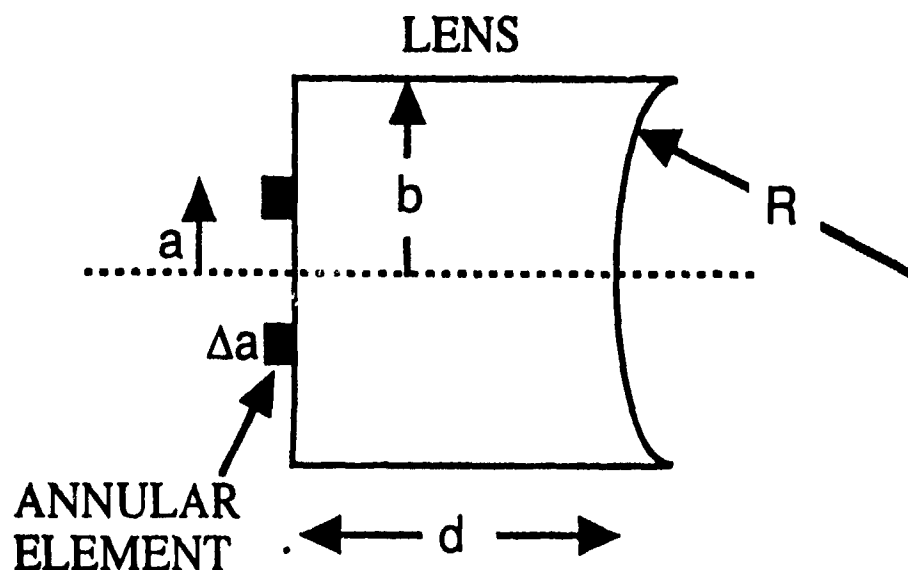


Figure 4. Proposed configuration for Bessel beam transducer.

REFERENCES

1. G. Du and M. A. Breazeale, J. Acoust. Soc. Am. 78, 2083-2086 (1985).
2. D. K. Hsu et al., IEEE Trans. Ultrason. Ferroelec. Freq. Contr. 37, 404-410 (1990).
3. J. Durnin, J.J. Miceli, and J.H. Eberly, Phys. Rev. Lett. 58, 1499-1501 (1987).
4. J. Durnin, J. Opt. Soc. Am. 4, 651-654 (1987).
5. R.W. Ziolkowski, D.K. Lewis, and B.D. Cook, Phys Rev. Lett. 62, 147-150 (1989).
6. D.K. Hsu, F.J. Margetan, and D.O. Thompson, Appl. Phys. Lett. 55, 2066-2068 (1989).
7. J.Y. Lu and J.F. Greenleaf, IEEE Trans. Ultrason. Ferroelec. Freq. Contr. 37, 438-447 (1990).

II. On-Axis Formulas for Ultrasonic Transducer Beams in Anisotropic Media

INTRODUCTION

The requirements of nondestructively inspecting cast steels, weldments, composites, and other inherently anisotropic materials have necessitated research aimed at better understanding ultrasonic wave propagation in these materials. Theoretical modeling can play an important role in the study of the unique effects, such as beam skew and excess beam divergence, which anisotropic materials have on ultrasonic transducer beams. The predictions of valid models can be used to help interpret nondestructive measurements and to help guide the choice of nondestructive inspection procedures by identifying configurations with the most promise of success. Much recent work has been devoted to this subject (for example Refs. [1-5]).

The purpose of the present work is to develop simple approximate formulas to predict the on-axis beam profile for ultrasonic transducers inspecting anisotropic materials. Many wave scattering theories assume plane wave incidence. These formulas can then be used, for example, to determine an incident wave amplitude for use with the scattering theories for predicting flaw response.

This approach has been used successfully in isotropic materials [6].

MODEL

The model configuration is shown in Fig. 1. A circular transducer of radius "a" and possibly focussed with focal length F is assumed to be insonifying a planar surface of an anisotropic solid from a water bath with an incident angle θ_i and standoff distance "d". The refracted phase velocity, V_p , is in the z-direction and has angle θ_r . The energy velocity, V_g , may have a skew angle ψ due to the anisotropy. Since the center of the beam energy will travel in the V_g direction, this will be defined as the beam axis, or z_{skew} . The only restriction will be that the plane of Fig. 1 must be a material symmetry plane.

The propagation of a time harmonic ultrasonic beam can be expressed in the form of an angular spectrum of plane waves, ASPW; i.e., a superposition of plane waves traveling in all directions, each with a particular amplitude

$$u(x,y,z) = e^{i\omega t} \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(k_x, k_y) e^{-i(k_x x + k_y y + k_z z)} dk_x dk_y \quad (1)$$

where ϕ is the 2-D spatial Fourier transform of the sound generating source; in this case, a planar or focussed piston. A detailed discussion of the ASPW formulation for anisotropic materials is found in Ref. [4]. The key to this formulation is k_z , the component of the wave vector in the z-direction (the direction of phase propagation). Given k_x and k_y as the integration variables, k_z is found via the slowness surface of the material, which is derived from the elastic constants.

The ASPW formulation lends itself well to treating transmission of the beam through a plane interface since refraction of plane waves at a plane interface is well understood. Each plane wave in the superposition is analyzed as it passes through the interface and the resulting spectrum of refracted waves is recombined to yield the field in the solid. Roberts has done this rigorously with a numerical integration of the resulting integrals for the case of a transversely isotropic solid [3].

The objective at present is to make some key approximations which will allow a more nearly closed form solution for the axial component of the transmitted beam. This follows the approach previously used for the isotropic case [7]. The fundamental approximation is the

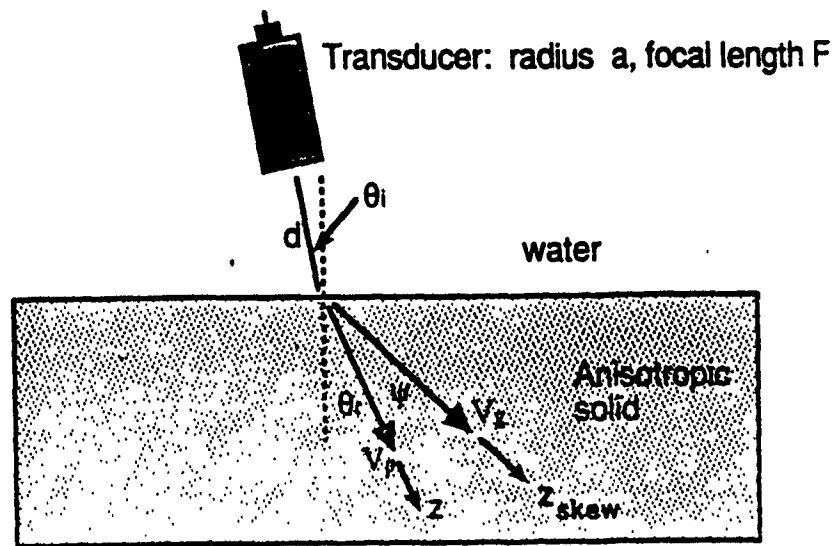


Figure 1. Model configuration.

paraxial assumption; i.e., the beam is assumed to be well collimated, with most of the energy in the angular spectrum concentrated in plane waves travelling in the vicinity of the axial direction. Several consequences follow. Provided that the inspection configuration is well away from any critical angles, the transmission coefficient will vary slowly over the small angular spectrum and can be assumed a constant.

More significantly, the value of k_z is, for an anisotropic material, generally not available in closed form. However, the paraxial approximation allows a local approximation to the slowness as a second order (parabolic) surface. the parameters obtained from this

parabolic fit are the phase velocity in the axial direction V_p , the skew angle ψ , and the curvatures Δ_x and Δ_y of the slowness surface in the x-z and y-z planes (plane of Fig. 1 and the plane orthogonal to V_p). This parabolic approximation is detailed in Refs. [4,8,9]. The curvature terms control the rate of divergence (diffraction spread) of the beam.

After the procedure outlined above has been carried out and the approximations employed, then the majority of the resulting integrals may be evaluated exactly for the axial case and the amplitude of the displacement field along the z_{skew} -axis can be written as

$$u = e^{j\omega t} e^{-j(k_0 d + k_1 z)} C_s(S, \Delta S) \quad (2)$$

where k_0 is the wave number in the water, k_1 is the wave number associated with the V_p direction in the solid and $C_s(S, \Delta S)$ is given by the integral

$$C_s(S, \Delta S) = \frac{\sqrt{S^2 - \Delta S^2}}{2\pi} \int_0^{2\pi} \left\{ \frac{1 - e^{-j\pi(S - R + \Delta S \cos 2\theta)/(S^2 - \Delta S^2)}}{S - R + \Delta S \cos 2\theta} \right\} e^{j\theta} d\theta \quad (3)$$

where

$$R = (S^2 - \Delta S^2) a^2 / \lambda_0 F \quad (4)$$

$$S = \frac{z_{\text{skew}} \lambda_1 \cos \psi}{2a^2} [\Delta_y + \gamma^2 \Delta_x] + \frac{d\lambda_0}{a^2} \quad (5)$$

$$\Delta S = \frac{z_{skew} \lambda_1 \cos \psi}{2a^2} [\Delta_y - \gamma^2 \Delta_x] \quad (6)$$

$$\gamma = \frac{\cos \psi \cos \theta_i}{\cos(\theta_r - \psi)} \quad (7)$$

The parameters S and ΔS can be thought of as generalizations of the familiar nondimensional quantity used in transducer theory to define the nearfield length: $S = z\lambda/a^2$. The parameter γ is a "refraction parameter" related to the change in beam width due to refraction at the interface.

The integral in equation (3) is well behaved and can be easily evaluated with a call to a computer math library integration routine. This integral may also be evaluated exactly by expanding the denominator of the integrand in a Fourier series and integrating term by term to yield a series representation of the solution

$$C_s(S, \Delta S) = \frac{\sqrt{S^2 - \Delta S^2}}{\sqrt{(S-R)^2 - \Delta S^2}} \left\{ 1 - e^{-j\pi(S-R)/(S^2 - \Delta S^2)} \sum_{n=0}^{\infty} b_n J_n \left(\frac{\pi \Delta S}{S^2 - \Delta S^2} \right) \right\} \quad (8)$$

where

$$b_0 = 1, \quad b_n = 2(-j)^n \left\{ \frac{\sqrt{(S-R)^2 - \Delta S^2} - S + R}{\Delta S} \right\}^n, \quad n = 1, 2, \dots \quad (9)$$

The solution may be computed from equation (8).

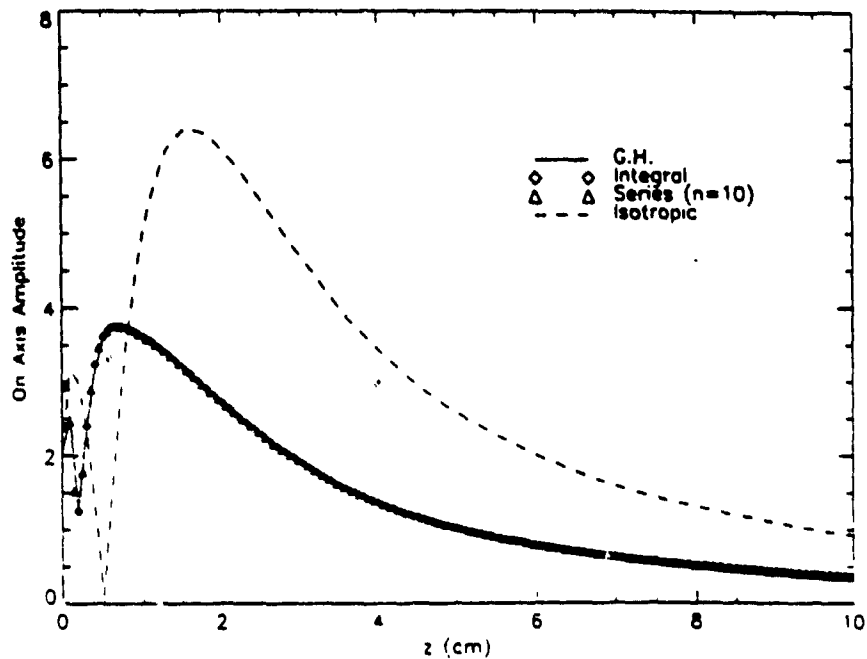
Convergence is generally achieved with less than 10 terms and more often less than 5 terms.

EXAMPLE CALCULATIONS

The on-axis formulas, both in integral form and series form, have been used to predict the axial profiles for ultrasonic inspection of a unidirectional graphite epoxy laminate. These predictions are compared with the predictions of the Gauss-Hermite beam model. The Gauss-Hermite beam model is a model which predicts the full-field within the paraxial approximation and which has undergone extensive experimental validation for cases including the one used here [4,8,9]. In order to illustrate the effects of the anisotropy, the axial profile which would be obtained in an isotropic material with the same nominal wavespeed are also shown on the figure.

Figure 2 shows the axial profile for a normal incidence inspection of a unidirectional composite using a focussed probe. The series and integral computations are virtually identical with each other and with the Gauss-Hermite solution. One significant result is the reduction of the peak amplitude of the beam at the focal point as

compared to the isotropic material. This is due to the fact that the anisotropy causes the focussing parallel and perpendicular to the fibers to occur at different depths, thus spreading out the focal region.



- Normal incidence to unidirectional graphite-epoxy panel
- Focussed probe, $a=.635$ cm, $F=7.62$ cm, $d=2.54$ cm, frequency=5 MHz
- Wavespeed in solid = $.313$ cm/ μ s, $\Delta x=3.56$, $\Delta y=0$, $\gamma=0$

Figure 2. Axial profile in unidirectional composite.

CONCLUSIONS

A simple formalism, easy to numerically implement, has been derived for the prediction of the on-axis beam profile of an ultrasonic transducer inspecting an anisotropic material. The results of this beam model agree with those of the more general Gauss-Hermite beam model. These results can be used to better understand the effects that anisotropy has on ultrasonic inspections and to help interpret measurement data from anisotropic materials.

REFERENCES

1. J.A. Ogilvy, *Ultrasonics* 24, 337-347 (1986).
2. A.N. Norris, *Wave Motion* 9, 1-24 (1987).
3. R.A. Roberts, *Ultrasonics* 26, 139-147 (1988).
4. B.P. Newberry and R.B. Thompson, *J. Acoust. Soc. Am.* 85, 2290-2298 (1989).
5. J.L. Rose, K. Balasubramaniam, A. Tverdokhlebov, *JNDE* 8, 165-179 (1989).
6. R.B. Thompson and T.A. Gray, *J. Acoust. Soc. Am.* 74, 1279-1290 (1983).
7. R.B. Thompson and T.A. Gray, in Review of Progress in Quantitative NDE, edited by D.O. Thompson and D.E. Chimenti (Plenum Press, New York, 1983) Vol. 2A, p. 567.
8. E.P. Papadakis, F. Margetan, and B.P. Newberry, in Review of Progress in Quantitative NDE, edited by D.O. Thompson and D.E. Chimenti (Plenum Press, New York, 1990) Vol. 9A, p. 235.
9. F.J. Margetan, R.B. Thompson, and T.A. Gray, in Review of Progress in Quantitative NDE, edited by D.O. Thompson and D.E. Chimenti (Plenum Press, New York, 1991) Vol. 10A, p. 1539.

POLYMER CONDUCTIVITY AND OPTICS

Joanne Promislow

ABSTRACT

Using a Model 110 Keithley Hall Effect Measurement System and a Model LTS-22 RMC cryogenic refrigeration system, preliminary photoconductivity data were obtained for a thin film sample of Kr implanted BBL. The cryogenic unit was upgraded to include a more evenly spaced heater coil and more accurate temperature sensors. Much effort was put into and some progress was made on optimizing the sample mounting procedure, so that valid conductivity research can be performed in the future. A new technique for determining the sample thickness and index of refraction of polymeric thin films from the reflectivity pattern of a P-wave was tested and found to be valid for isotropic thin films.

INTRODUCTION

Conductivity:

Conductive polymers have generated an enormous amount of interest and been the subject of unrelenting research during the past decade. Until recent times, polymers were classified strictly as insulators to the conduction of electric current because of their relatively large band gaps. In the late seventies, however, certain polymers were discovered to have surprisingly low resistivities (indicating high conductivities since conductivity is the reciprocal of resistivity) when properly doped. This discovery caused the

previously nonexistent field of electrically conducting polymers to explode into existence as polymers having conductivities in the semiconductor and conductor range were synthesized.

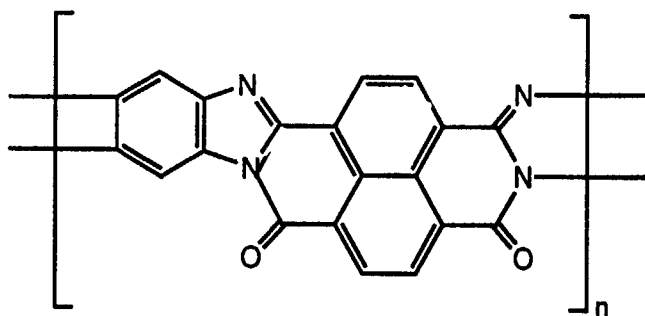
The abundance of applications for conductive polymers provides the primary motivation for exploring polymer conductivity. Electroconductive polymers are currently being used in rechargeable batteries and the development of plastic all solid-state batteries with energy densities several times that of typical batteries is envisioned. Other immediate applications are in areas such as static dissipation and microwave shielding and scattering. Radar and lower frequency microwave absorbers, which can increase or decrease radar cross-sections by their levels of conductance are certainly of interest to the military. In addition to these immediate applications, there are several others in mind for the future. For example, some researchers believe that highly conductive plastic wires will be an actuality by the turn of the century. The resulting weight savings would produce very significant industrial benefits, particularly in aeronautics. Other possible future applications for conductive polymers include use in fuel cells, capacitors, EMI shielding, biochemical sensors, and neutron detection.

Another reason for developing electro-conducting polymers which is more specifically related to aeronautics is for their potential role in eliminating spacecraft charging. Spacecraft charging refers to the accumulation of secondary electrons by polymer dielectrics used in spacecraft when exposed to the space environment. This accumulation results in a growing electrostatic potential, which, when it exceeds the dielectric strength of the

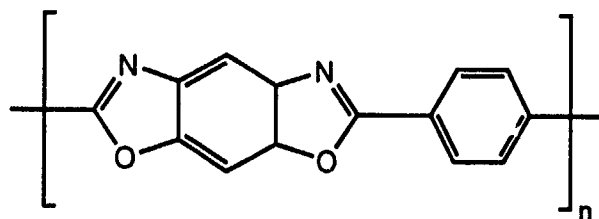
polymer, can cause a breakdown of the polymer and consequently seriously damage spacecraft functioning. By developing polymer dielectrics with a moderate electrical conductivity, this spacecraft charging problem could be overcome. The damaging charge would be conducted away faster than the electrons accumulated and thus discharge would be prevented. In addition to being sufficiently conductive, polymers used in this application would also have to meet other criteria necessary for use in space, such as mechanical strength and environmental stability. For this reason, many high performance polymers are currently being researched for potential use in this area.

For all of these reasons, polymer conductivity is currently a subject of intense research. The rush is on to develop highly conductive, environmentally stable, and processable, polymers. However, there are still many obstacles to overcome. Owing to the newness of conductive polymers, many basic aspects of the field, such as mechanisms of conductance, polymer formation, and decay processes, are still poorly understood. It was this need for a more thorough understanding of these fundamental processes which motivated the research I was involved in this summer at Wright Patterson Air Force Base. I participated in ongoing research of the conductive mechanisms of various highly ordered polymers. As previously mentioned, research of highly ordered polymers which possess exceptional mechanical properties and environmental stability is of particular interest to the Air Force because of the suitability of these polymers for aeronautical applications. Research

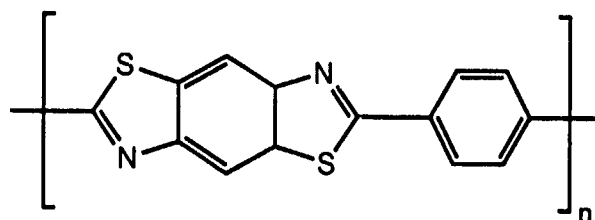
was therefore concentrated on polymers geared towards aeronautical use, such as BBL, PBO, PBT, and their derivatives.



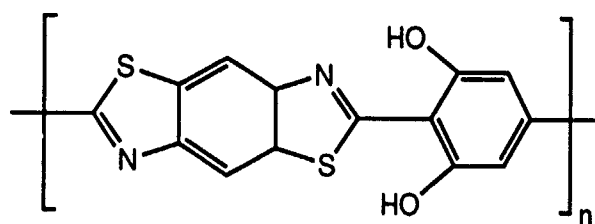
BBL



PBO



PBT



DPBT

Optics:

In addition to researching polymer conductivity, I was also involved in studies regarding the optical properties of polymeric thin films. Polymers play a key role in such optical phenomena as wave-guiding and non-linear optics, areas which are predicted to revolutionize technology in the near future. To do research in these areas, the linear optical properties, such as the sample thickness and index of refraction, of the sample to be studied must be known. Thus, determining the linear optical properties of polymeric thin films is a very necessary task. Unfortunately, it is often a tedious one as well, and involves making several measurements. The purpose of the optical research I was involved in this summer was to test a method for calculating both the sample thickness and index of refraction of polymeric thin films by studying their reflectivity interference patterns. The idea was that if these two pieces of information could be determined from just one relatively simple measurement, this would be a very valuable and convenient technique for optics research.

PROCEDURE

Conductivity:

The overall goal of the polymeric conductivity research I participated in was to determine the mechanism of conduction for several highly ordered polymers. In determining the conduction mechanism for a polymer, it is useful to have as much information

on the conductivity of that polymer as possible. The temperature dependence of the polymer's conductivity is of particular interest and most of my research efforts this summer were concentrated on obtaining this information. For this purpose a RMC cryogenic refrigeration system and a Keithley hall effect measurement system were employed. All measurements were taken on samples in the form of thin films. The thin films were mounted on a sample stage inside a vacuum chamber in the cryogenic unit. This set up allowed the sample to be studied over a temperature range of 320 K to approximately 12 K. The Keithley hall effect system was the computer controlled test system which was used to make sample resistivity measurements via the classic Van der Pauw technique. In the Van der Pauw technique, four electrical connections are made to the sample and as current is forced between two of the connections, the voltage is measured across the other two. Measurements are taken as the current is forced in both directions and across all combinations of the four electrical connections. Thus eight voltage and current values are obtained in total and by performing an averaging calculation on these values, the resistance, R , of the sample can be computed. Finally, if the thickness of the sample is known, the sample resistivity can easily be computed from the equation $\rho = R * L$, where ρ is resistivity, and L is the thickness. A plot of resistivity (usually expressed on a logarithmic scale) versus some power of temperature, T , (often the inverse temperature is used) is the conventional way to represent the temperature dependence.

To gain additional information, some photoconductivity experiments were performed on one of the samples. Over the entire temperature range, resistivity measurements were taken both while the sample was protected from irradiation and while the sample was being exposed to a monochromatic light source. By placing neutral density filters in front of the light source, the intensity of the irradiation could be controlled and conductivity could thus be measured as both a function of the intensity of irradiation and temperature. These experiments were performed using a helium-neon laser ($\lambda=632$ nm) and an adjustable tungsten light source which was tuned to a wavelength of 1000 nm.

Optics:

Polymeric thin films were also used for the optical measurements. To measure the refractivity pattern of a sample an ellipsometer was used, with the light source polarized so that the electromagnetic wave was parallel to the "plane of incident" (i.e. the P-wave was used). The interference pattern of the sample could then be acquired by using a power meter to measure the intensity of the refracted beam as the sample was rotated through the desired range of reflective angles. Due to machine limitations, values could only be obtained for reflective angles over the range of 85° to 23° . Measurements were made on both isotropic and anisotropic samples of Mylar.

RESULTS

Conductivity:

The first sample which was studied was a Kr (0.19MeV) implanted BBL thin film. The resistivity vs. temperature curve was first measured with the sample in the nonirradiated state. The resistivity measurements were then repeated over the same temperature range for the sample under three different conditions of irradiation. In one run the sample was exposed to the full intensity of the helium neon ($\lambda=632$ nm) laser beam. In another the sample was exposed to the laser beam after it had been filtered to 77% of its original power and in the last the sample was exposed to the laser beam after it had been filtered to 39% of its original power. All measurements were taken continuously over the temperature range for each given irradiation condition. The results are shown in plot 1 of the Appendix.

Although the fact that the non-irradiated sample consistently had the highest resistivity values does give some indication of photoconductivity, the data were observed to overlap a lot. It was decided that laser drift from day to day could be responsible for this overlap. For this reason the experiment was repeated but this time instead of continuously measuring over the entire temperature range for each given irradiation condition and then switching to the next set up and repeating the temperature scan, only one sweep over the temperature range was made. At each temperature, the resistivity measurements for all four of the irradiation conditions were taken before moving on to the next temperature. This was done by covering up the sample holder windows for the non-

irradiated measurement, placing the appropriate neutral density filter in front of the laser for the filtered measurements, and letting the unfiltered laser beam shine directly on the sample for the full intensity measurement. Approximately two to three minutes were allowed between each measurement to give the sample sufficient time to adjust to the new condition. The results of this experiment are shown in plot 2 of the Appendix. As can be seen, the spread between the curves is somewhat clearer in this plot and it does appear that some photoconductivity is occurring.

The experiment was then repeated using a tungsten light source tuned to a wavelength of 1000 nm instead of the helium neon laser. The results of this experiment are shown in plot 3 of the Appendix. Again, despite some overlap between the curves, there is some evidence of photoconductivity. What was unsettling about the plot, however, was that the resistivity values for the non-irradiated sample were observed to be significantly lower (a factor of 100) than they had been for previous experiments. We tested for long term excitation effects by irradiating the sample for several hours, taking a few resistivity measurements, and then turning the light source off and continuously taking measurements until the resistivity values levelled off. No long term excitation was observed, however, since this time the resistivity values simply jumped back up to their normal level and stayed there as soon as the light source was blocked. This finding led us to question the consistency of our measurement system. Plotting all of the non-irradiated measurements that had been made over the past several weeks on

the same graph confirmed our suspicions that a problem existed. This graph is shown in plot 4 of the Appendix.

To test the measurement system we took conductivity measurements on the sample at a fixed temperature, but over a range of forced currents. We found that at lower temperatures, when the measured voltage and current values were plotted, the data were not linear as they should have been. An example of this problem is shown in plot 5 of the Appendix which plots the voltage and current readings obtained at 20 K. Also, we noticed that even at higher temperatures, for a given forced current, the voltage and current measurements were not very consistent. When we opened up the cryogenic unit to look for problems and change the sample, we noticed that the wiring on the heating coil was not very evenly spaced and that there was a large burn spot on one area of the coil, which we suspected could be the source of our problems. The cryogenic unit was subsequently sent away to get this problem repaired. In addition to having the heating coil rewired, we replaced both of the temperature sensors with much more sensitive models. The new sample thermometer was also smaller and thus could be mounted closer to the sample.

While we were waiting for the cryostat to be repaired, we ran some room temperature conductivity measurements on a sample of dihydroxy PBT (DPBT) which had been previously irradiated with gamma rays. We found the room temperature conductivity to be 4.8×10^{-9} (S/cm). This value indicated that some long term radiation induced conductivity existed because even though it was a factor of 10 less than the value obtained immediately after

irradiation, it was significantly larger than the value obtained for the non-irradiated DPBT.

While waiting for the cryostat to be repaired, we continued to test our measurement system by measuring the resistivities of a one megohm and a one gigaohm resistor over a range of forced currents at a single temperature. We found that the accuracy of the machine was good and that much better precision was obtained at higher forced currents. Consequently, the remainder of the experiments run on the Keithley instrument were run at higher forced currents.

When the cryogenic unit was returned, before performing more polymer research, we first wanted to test it to ensure it was functioning properly. To do this a semiconductor sample of gallium-doped silicon with a known resistivity vs. temperature curve was obtained from our Electromagnetic Materials Division (MLP) and mounted on the sample stage. When we measured the resistivity of the sample over a temperature range of 320 K to 15 K and compared our results to the calibration curve, we found that our values agreed well with the calibration curve down to about 100 K, but that at colder temperatures our values were consistently lower and eventually levelled off. These results are shown in plot 6 in the Appendix. These results seemed to indicate that we had poor thermal contact, and that the sample was not getting as cold as the temperature sensors indicated. One good thing about the data however was that the curve we obtained looked very smooth and at each temperature the eight measured voltages and currents were very consistent.

In an attempt to remedy the thermal contact problem, several changes were made in mounting the sample. The sample was washed in warm trichloroethylene, then in warm acetone, and finally in methanol. We replaced the sapphire substrate plate we had been using with a beryllium oxide substrate plate which we had cleaned in trichloroethylene using an ultrasonic cleaner. Probably the most significant change made was the removal of a dental floss tie which was found in between layers of the sample stage and which was definitely interfering with good thermal contact. Presumably, the tie had been accidentally left on by the people who had mounted the temperature sensor. Finally, instead of using vacuum grease to hold the sample to the substrate plate, we placed a thin sheet of Teflon on the sample and a piece of rubber on top of the Teflon and used a copper tie to hold the sample to the substrate plate. We did this because we thought the vacuum grease might be conductive at the lower temperatures.

After making all of these changes, the experiment was rerun. This time the resistivity values which were obtained agreed excellently with the calibration curve, as is shown in plot 7 of the Appendix. The only problem was that the experiment stopped running at 27 K, indicating that one of the electrical connections to the sample had probably become disconnected. The unit was opened up and the loose connection resoldered. However, when the experiment was run again, several problems arose. At low temperatures, there was a significant discrepancy between the readings on the two temperature sensors. Also, the data that were obtained no longer agreed well with the calibration curve. They

levelled off at low temperatures and there was a big jump in the resistivity values between 40 K and 35 K. These results are shown in plot 8 of the Appendix. The cryogenic unit was opened up again but no problems could be spotted, except for a rather unsettling burning smell, so the unit was closed back up and the test was rerun. Again, the values obtained were not in good agreement with the calibration curve, but this time the data were too high instead of too low. These results can be seen in plot 9 of the Appendix. Thus not only were the results inaccurate, they were inconsistently inaccurate. The cryogenic unit was reopened and it was found that the new temperature sensors may have been poorly connected. The temperature sensor was reconnected to the cryogenic unit using a special epoxy for cold temperatures. While waiting for the glue to dry, the calibration curve for the semiconductor sample was verified by the MLP scientists who owned it.

The resistivity versus temperature scan was performed twice more and as can be seen by plot 10 in the Appendix the plots still showed a large discontinuity at 40 K and did not match the calibration curve very well. This time, however, the two sets of data were at least consistent with each other.

Optics:

The first optical measurements were taken on a thin film of isotropic Mylar. The reflectivity pattern was measured three times: once along the machine direction, then with the sample rotated 45° from the machine direction, and finally perpendicular to the machine direction. As can be seen from plot 11 in the Appendix, in

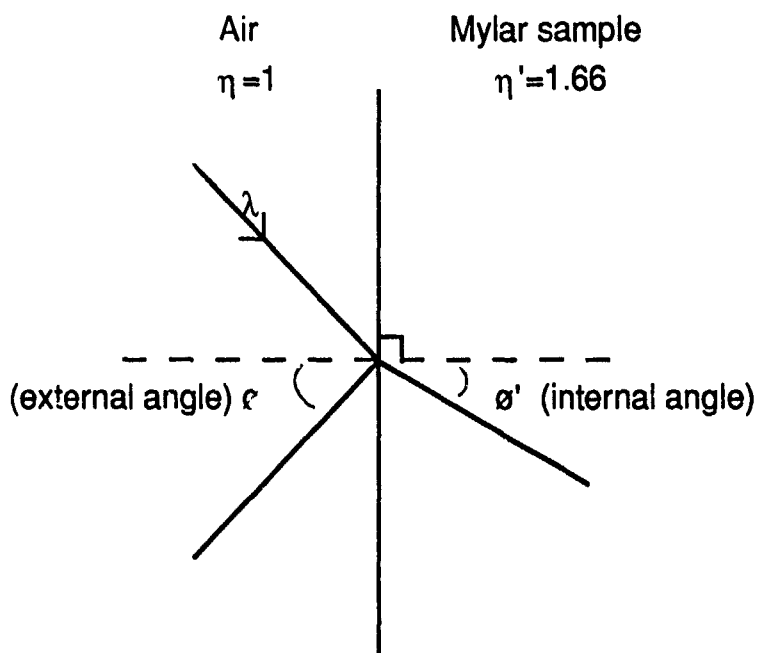
which reflectivity is plotted versus the reflective angle, the results from all three orientations are very consistent which verifies the isotropic nature of the sample. Plotting the reflectivity on a logarithmic scale (plot 12 in the Appendix) allows the minima to be more accurately seen, and the Brewster angle, which is the angle at which the reflectivity envelope is at an absolute minimum, to be determined. For the isotropic Mylar sample, the Brewster angle was determined to be 59° . From this value the index of refraction, η , of the sample can be easily calculated since $\eta = \tan \theta_B$, where θ_B is the Brewster angle. Thus for the isotropic sample the index of refraction was computed to be:

$$\eta = \tan (59^\circ) = 1.66$$

Once the index of refraction is known the sample thickness, d , can also be computed using the following formula:

$$d = \lambda / [2\eta' (\cos \phi'_A - \cos \phi'_B)]$$

where η' is the index of refraction of the sample, λ is the wavelength of the incident light (in this case 632 nm) and ϕ'_A and ϕ'_B represent the internal angles at which neighboring maximum or minimum reflectivity intensities occur.



To begin the calculation, one must first find the neighboring external angles at which either maximum or minimum reflection intensities are observed. For example, in the isotropic Mylar sample, it can be observed from the plot that the maximum reflectivity values were obtained at the external angles (θ_{\max}) shown in the chart below.

θ_{\max}	θ'_{\max}	$\cos \theta'_A - \cos \theta'_B$	
75°	35.6°		
65.7°	33.2°	0.227	
55.7°	29.8°	0.0320	
49.2°	27.1°	0.0224	
42.3°	23.9°	0.0240	
35.2°	20.3°	0.0236	
27.2°	16.0°	0.0234	
			avg=0.0247

These \emptyset_{\max} values can then be converted into the corresponding internal angles (\emptyset'_{\max}) using Snell's law:

$$\eta \sin \emptyset = \eta' \sin \emptyset' \quad \text{or}$$

$$\emptyset' = \sin^{-1} [(\eta \sin \emptyset)/\eta'] \quad \text{which in this case is}$$

$$\emptyset' = \sin^{-1} [\sin \emptyset/1.66]$$

Once these values are known, all the values for $\cos \emptyset'_A - \cos \emptyset'_B$ can be easily computed and the average can be taken to plug into the original equation:

$$d = \lambda / [2\eta' (\cos \emptyset'_A - \cos \emptyset'_B)]$$

Making the appropriate substitutions yields:

$$d = 632 * 10^{-9} / [2 (1.66) (.0247)]$$

$$d = 7.71 \mu\text{m}$$

This result agrees fairly well with the value obtained from spectroscopic measurements which was $6.8 \mu\text{m}$. The percent error between the two values is 12%.

Optical measurements were also taken on a thin film of anisotropic Mylar. Again measurements were taken for three orientations of the sample: along the machine direction, perpendicular to the machine direction, and at a 45° angle to the machine direction. The plot of reflectivity vs. reflective angle for each of the three measurements is shown in plot 13 of the Appendix. In plot 14, the same plot is shown on a logarithmic scale to emphasize the minima. As can be seen, this time the data are different for each orientation, as would be expected for an anisotropic sample. Although the Brewster angle can still be determined for each set of data by observing at what angle the minima of the interference envelope occurs, calculating the sample

thickness is no longer so simple. Because the sample is anisotropic, the index of refraction varies with orientation and an equation which deals with this complication has not yet been developed.

CONCLUSIONS

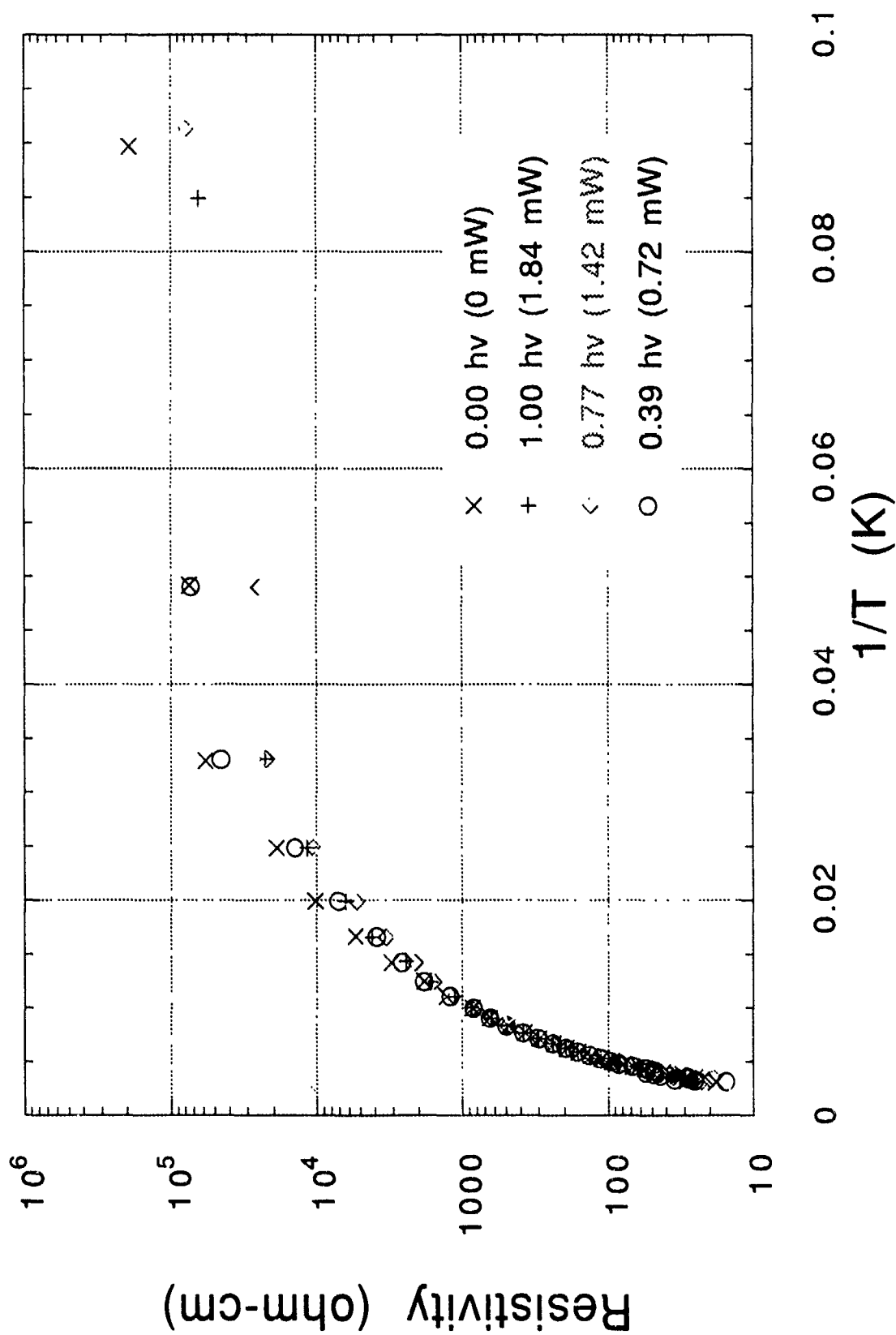
Although not too much headway has been made towards determining the conductive mechanisms of polymers, a great deal of effort was put into troubleshooting the measurement system and procedure. In that respect, a lot of progress was made towards identifying and eliminating the bugs in the system so that valid conductivity research can be made in the future.

Comparatively, the optics research proceeded much more smoothly. The data obtained for the interference pattern of the reflected P-wave were quite reasonable and the calculations performed on the data for the isotropic Mylar sample yielded fairly accurate values for the Brewster angle, index of refraction, and sample thickness. The only drawback to this technique is that it cannot yet be applied to anisotropic films and many of the polymers being most actively researched today are ordered. However, if this obstacle is overcome, the relative ease of this measurement should make it a convenient and useful technique in optics research.

APPENDIX

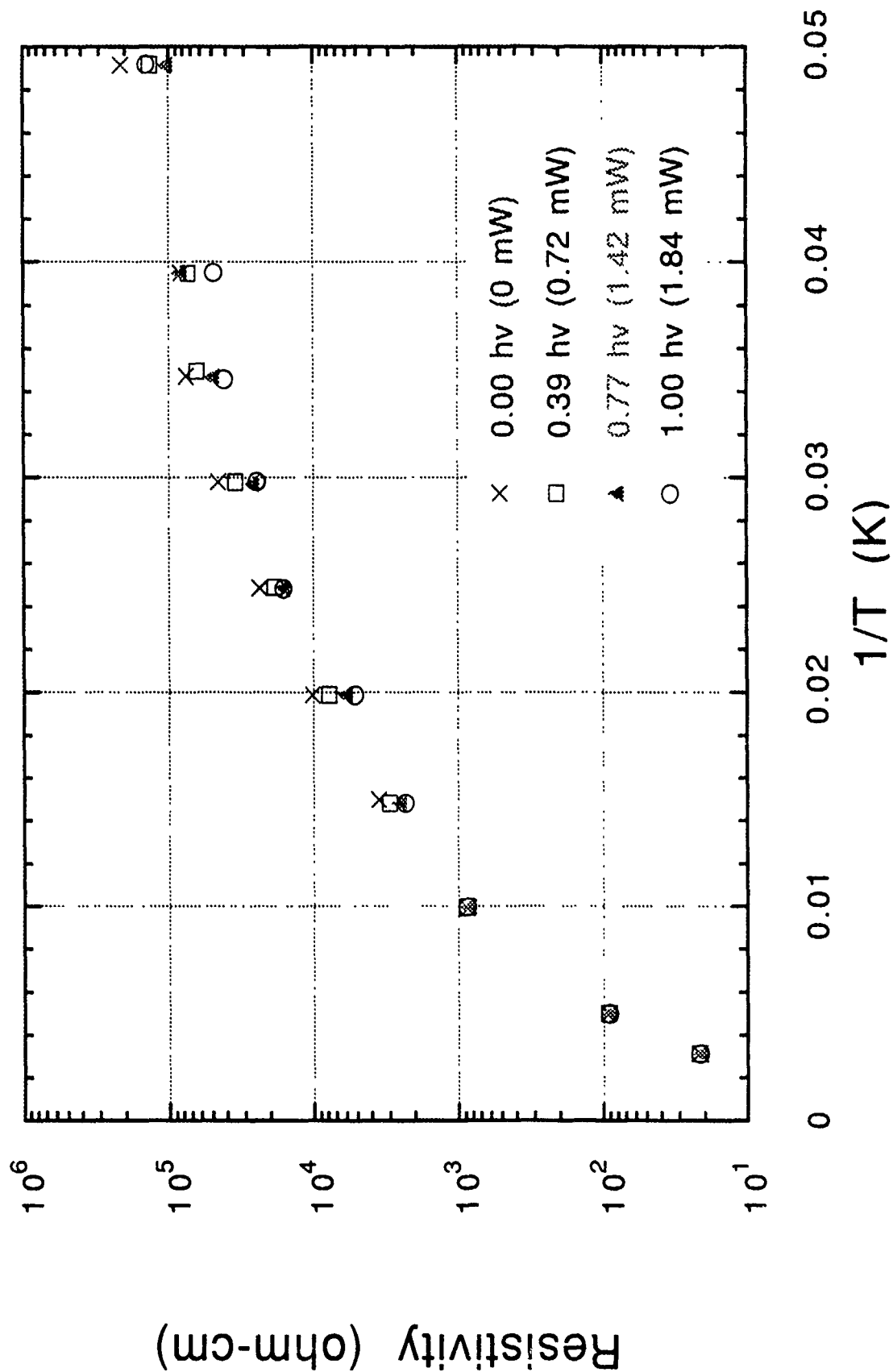
Plot 1

Kr (0.2 MeV) Irradiated BBL Film



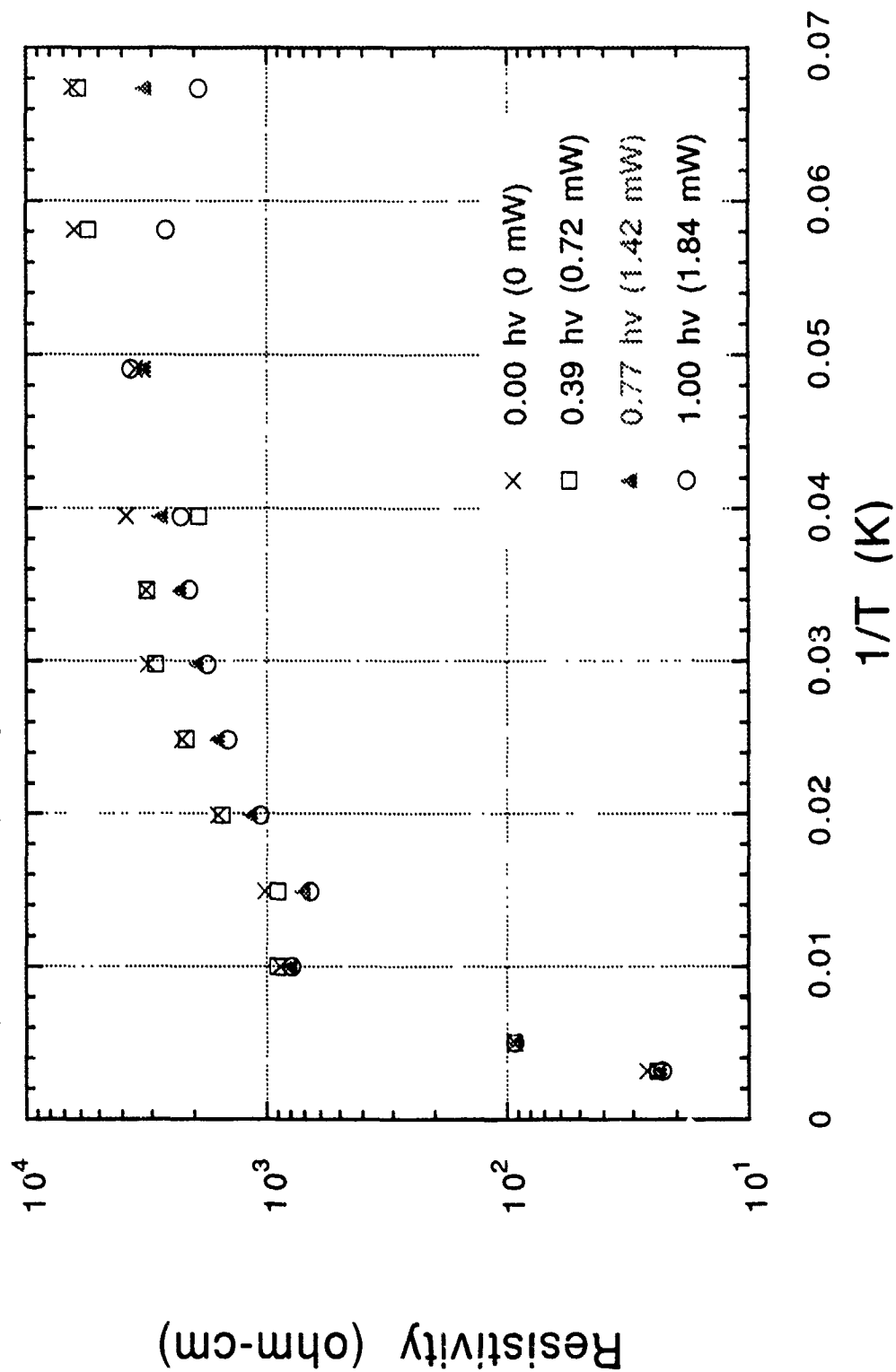
Plot 2

Kr (0.2 MeV) Implanted BBL Film



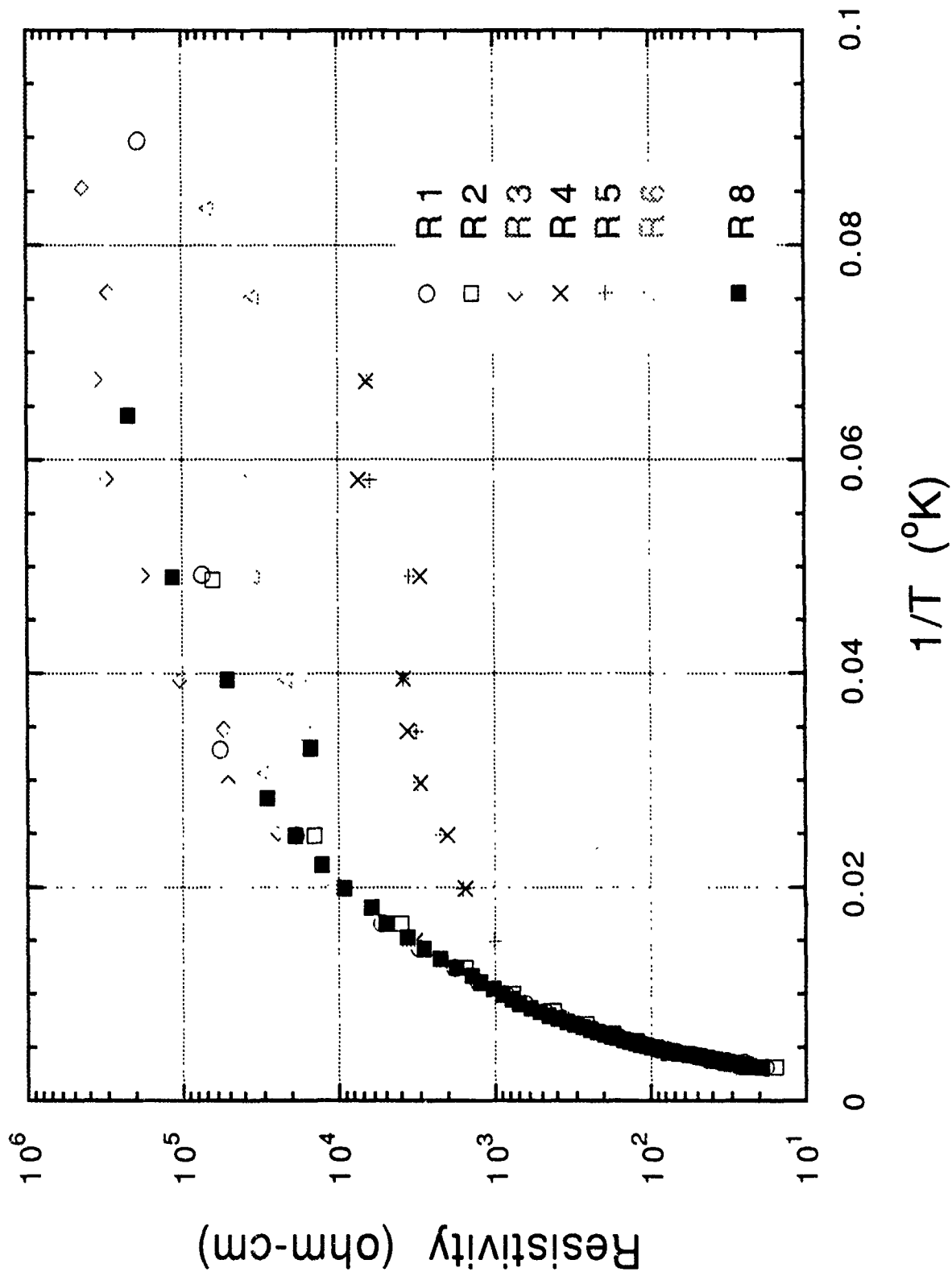
Plot 3

Kr (0.2 MeV) Implanted BBL film ($\lambda=1000\text{nm}$)

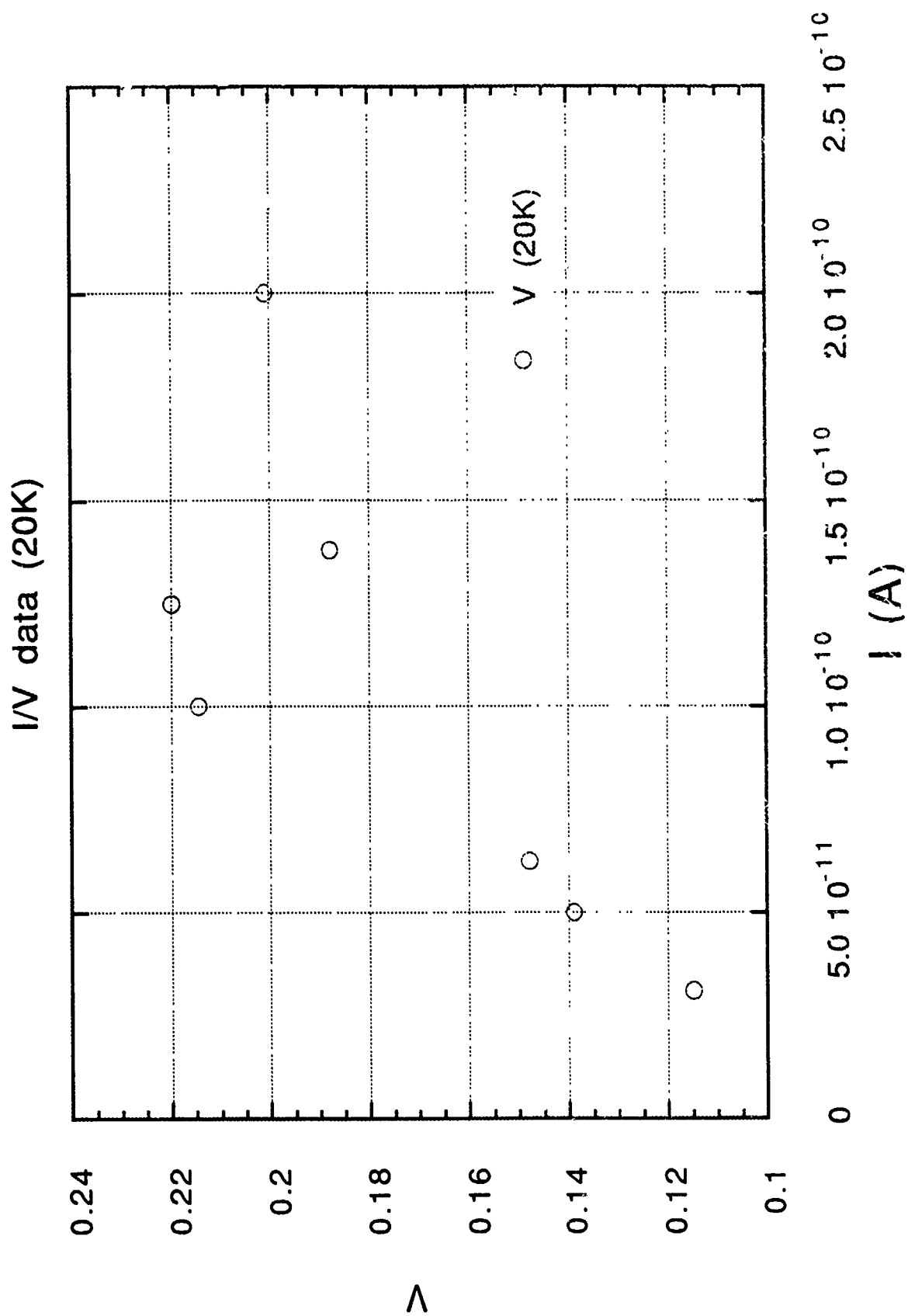


Plot 4

Reproducibility of no-hv meas. (BBL)

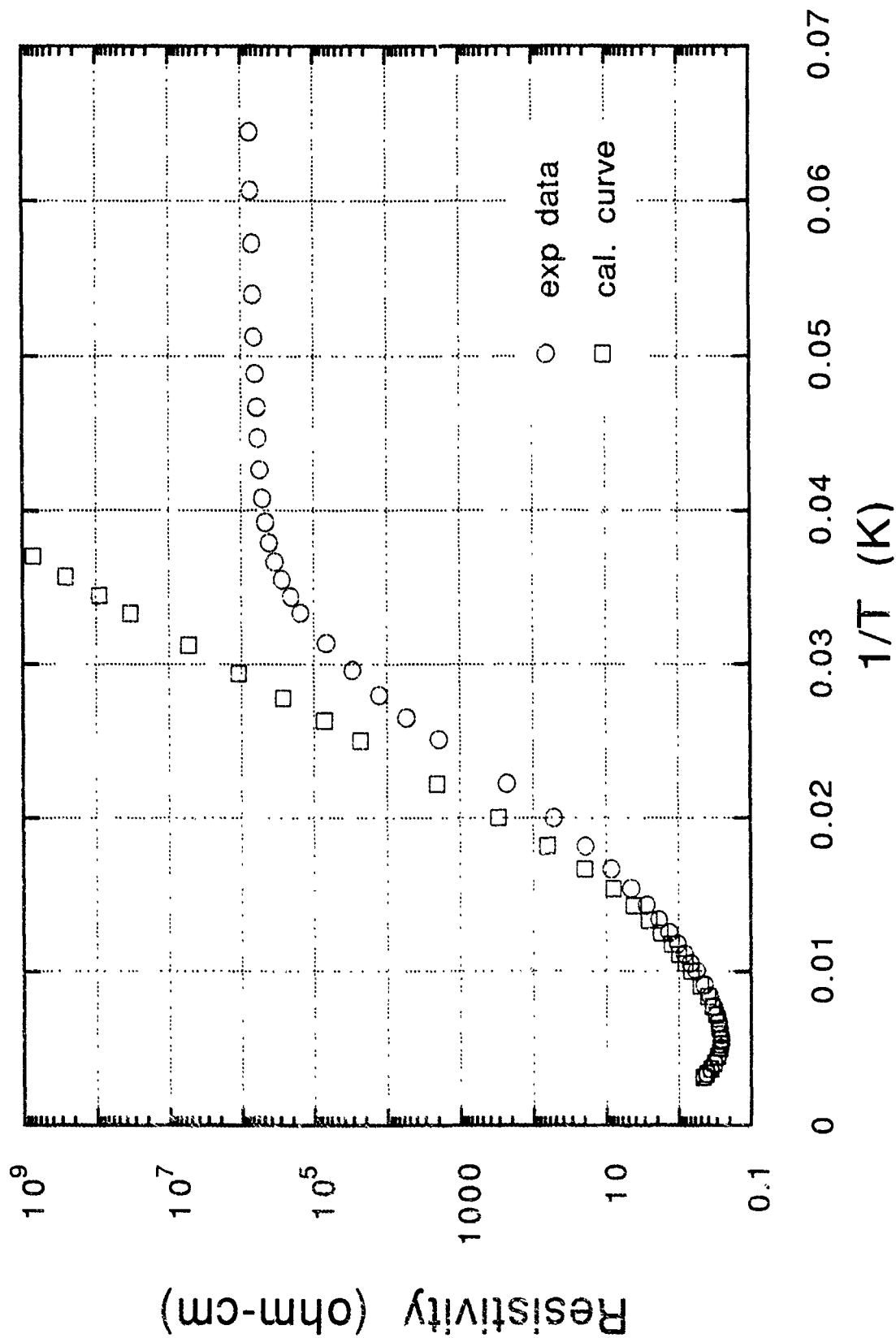


Plot 5



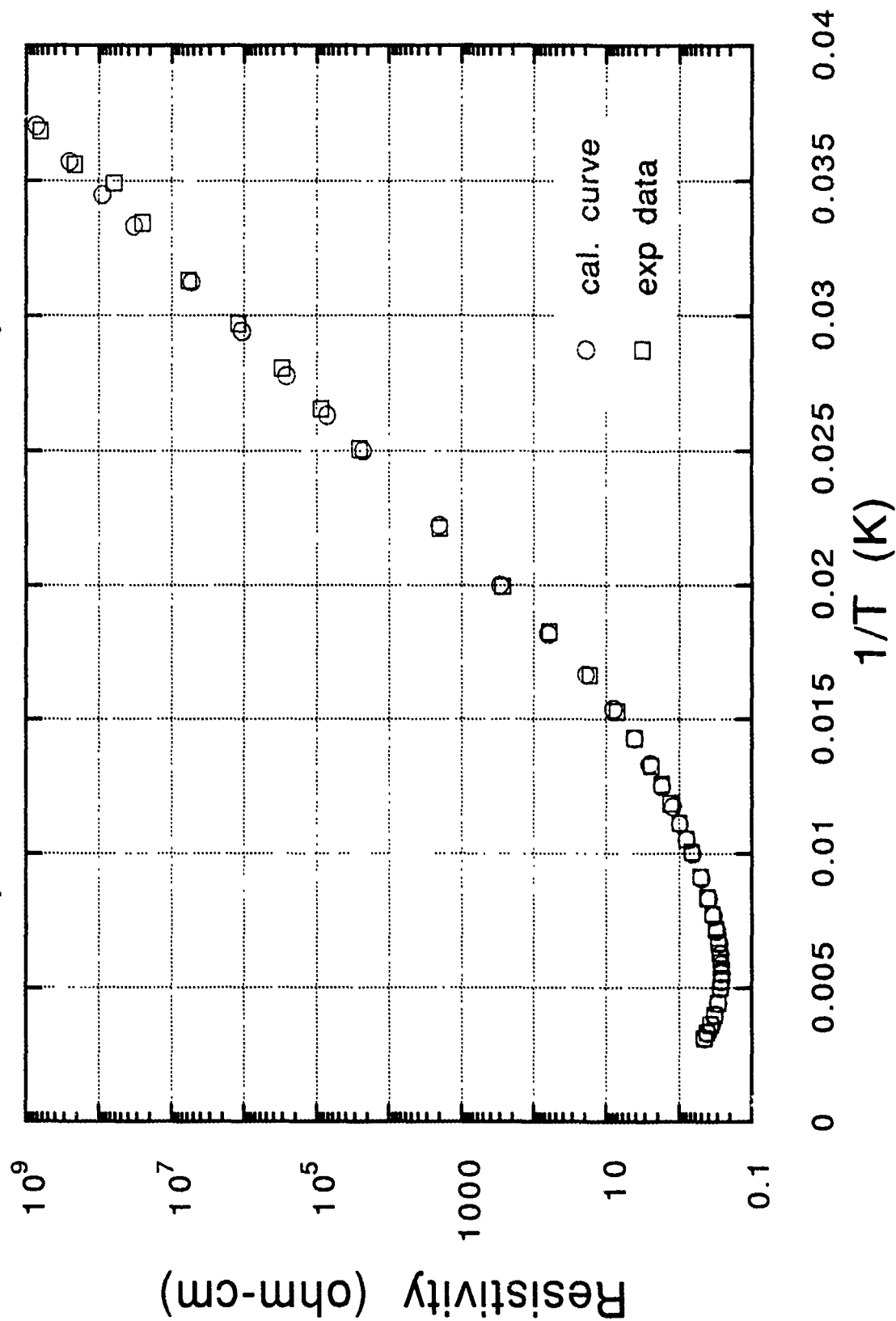
Plot 6

ρ vs. $1/T$ for semiconductor sample



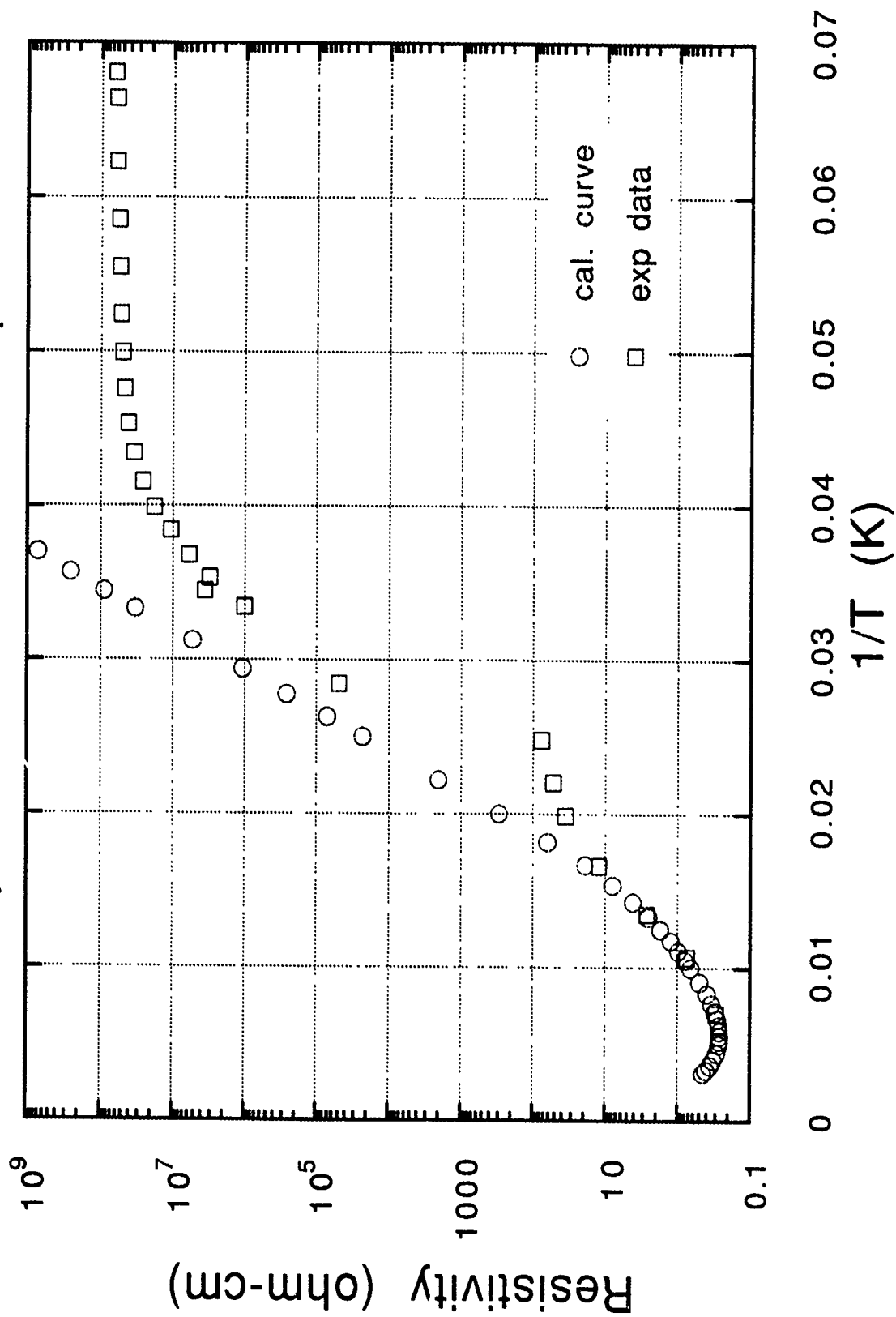
Plot 7

ρ vs. $1/T$ for Si: Ga sample



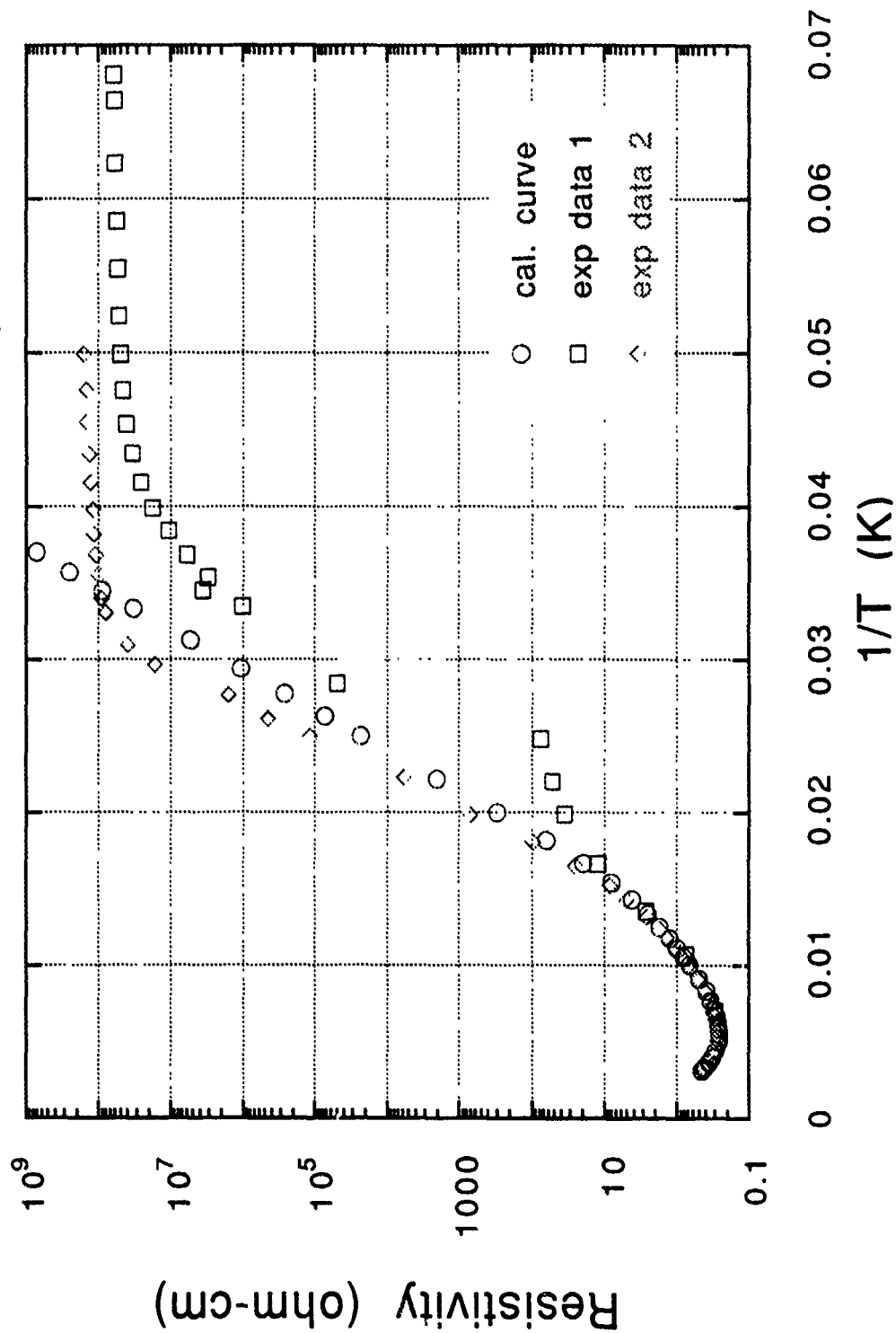
Plot 8

ρ vs. $1/T$ for Si: Ga sample



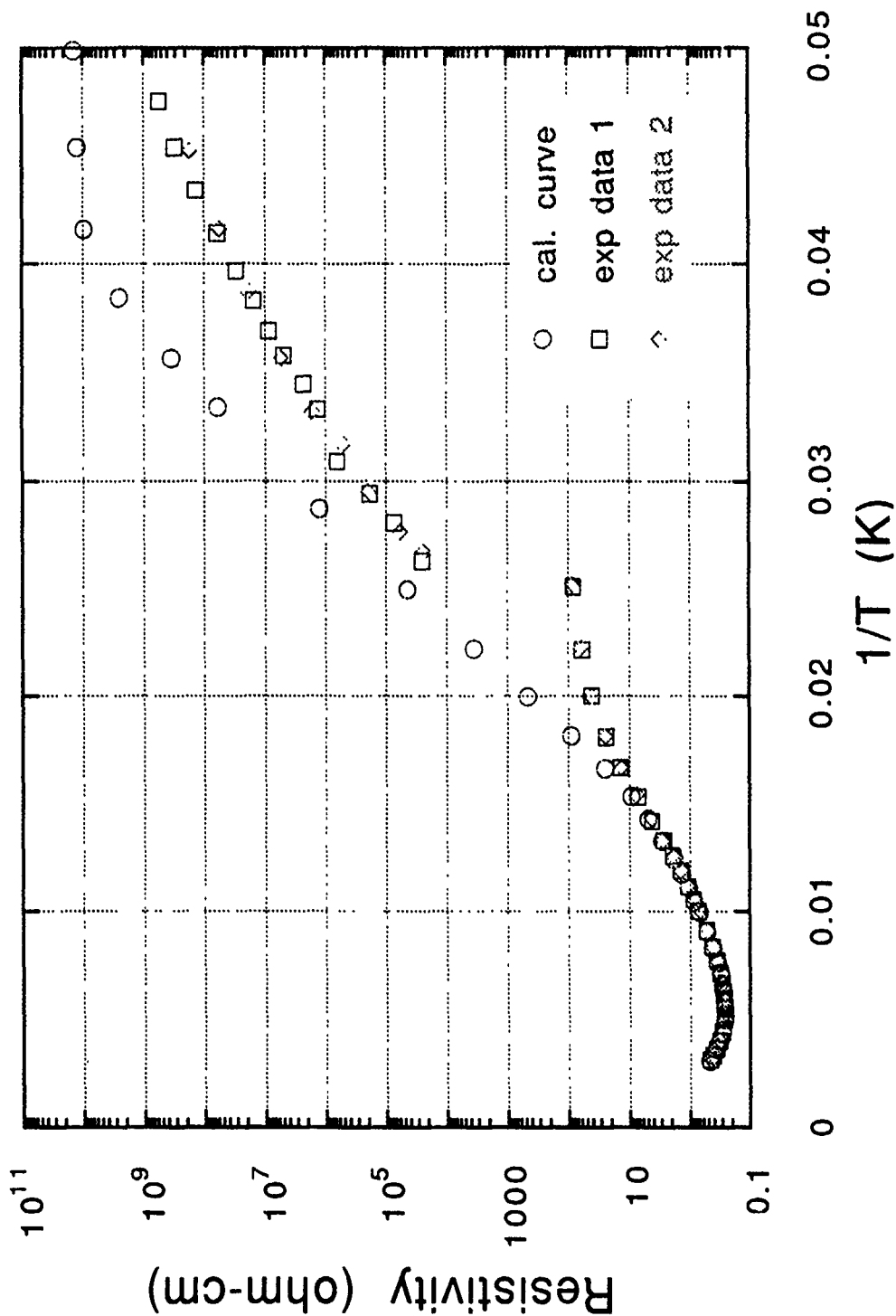
Plot 9

ρ vs. $1/T$ for Si: Ga sample



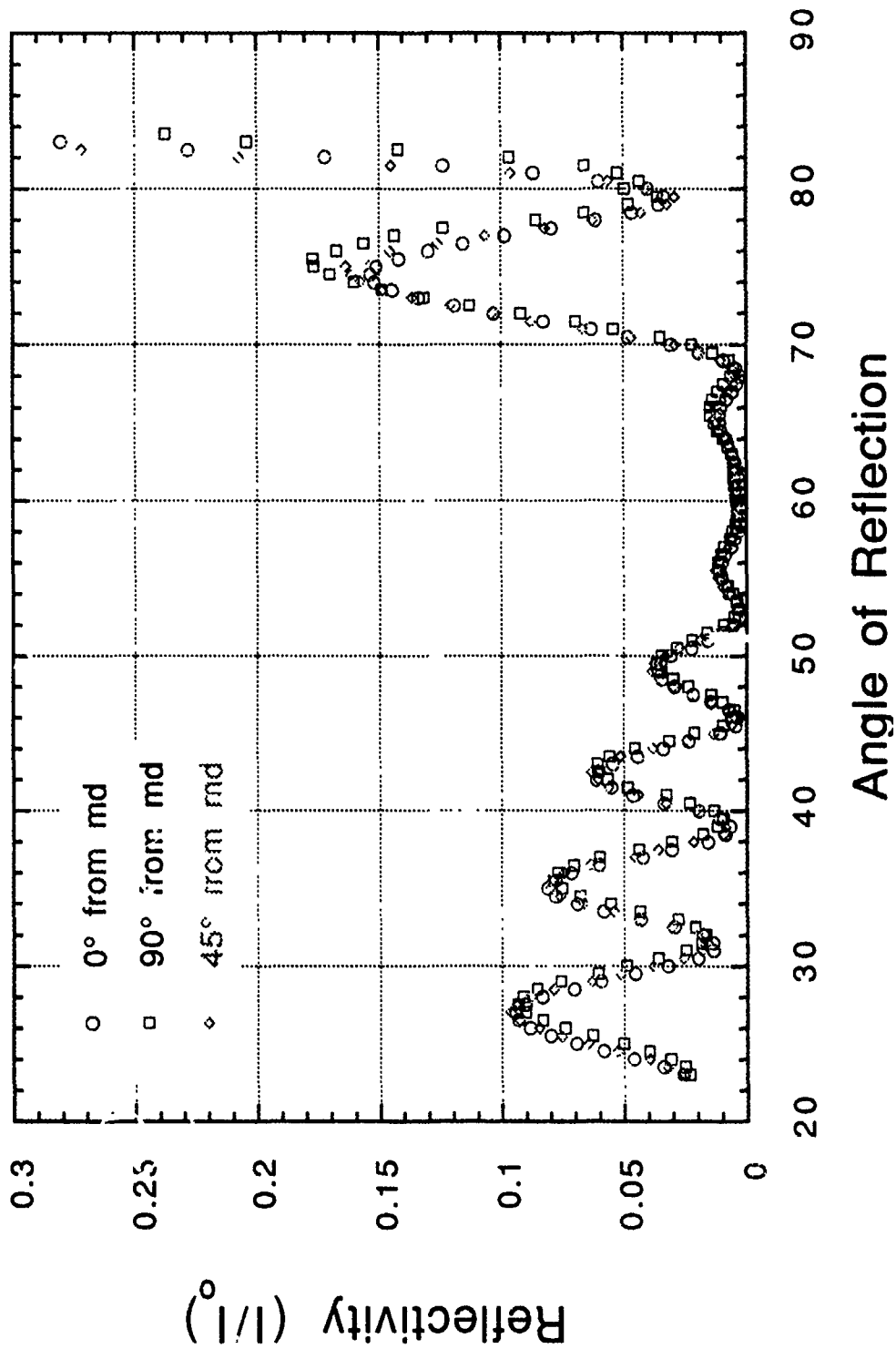
Plot 10

ρ vs. $1/T$ for Si: Ga sample



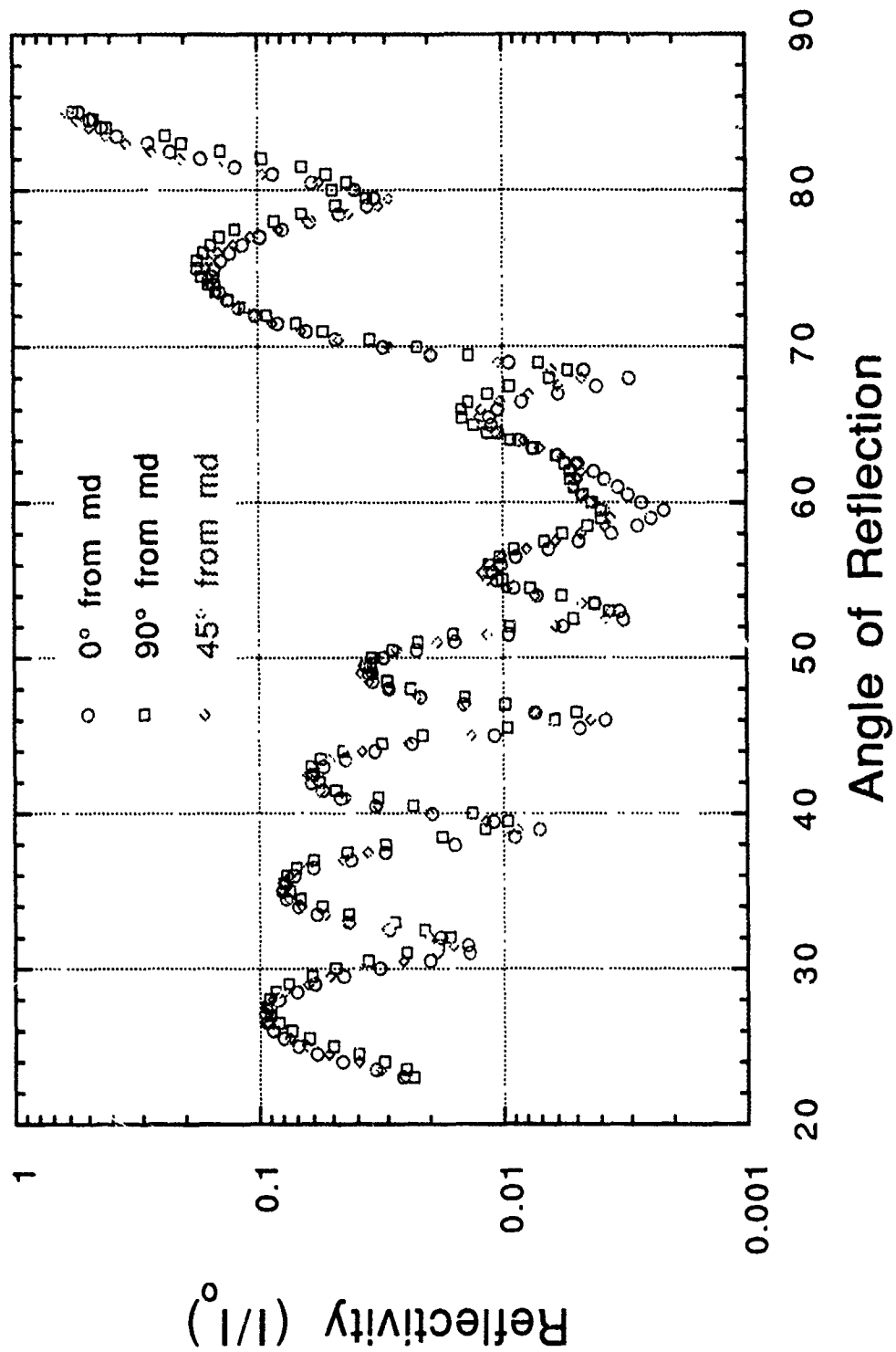
Plot 11

I vs. Θ : isotropic Mylar



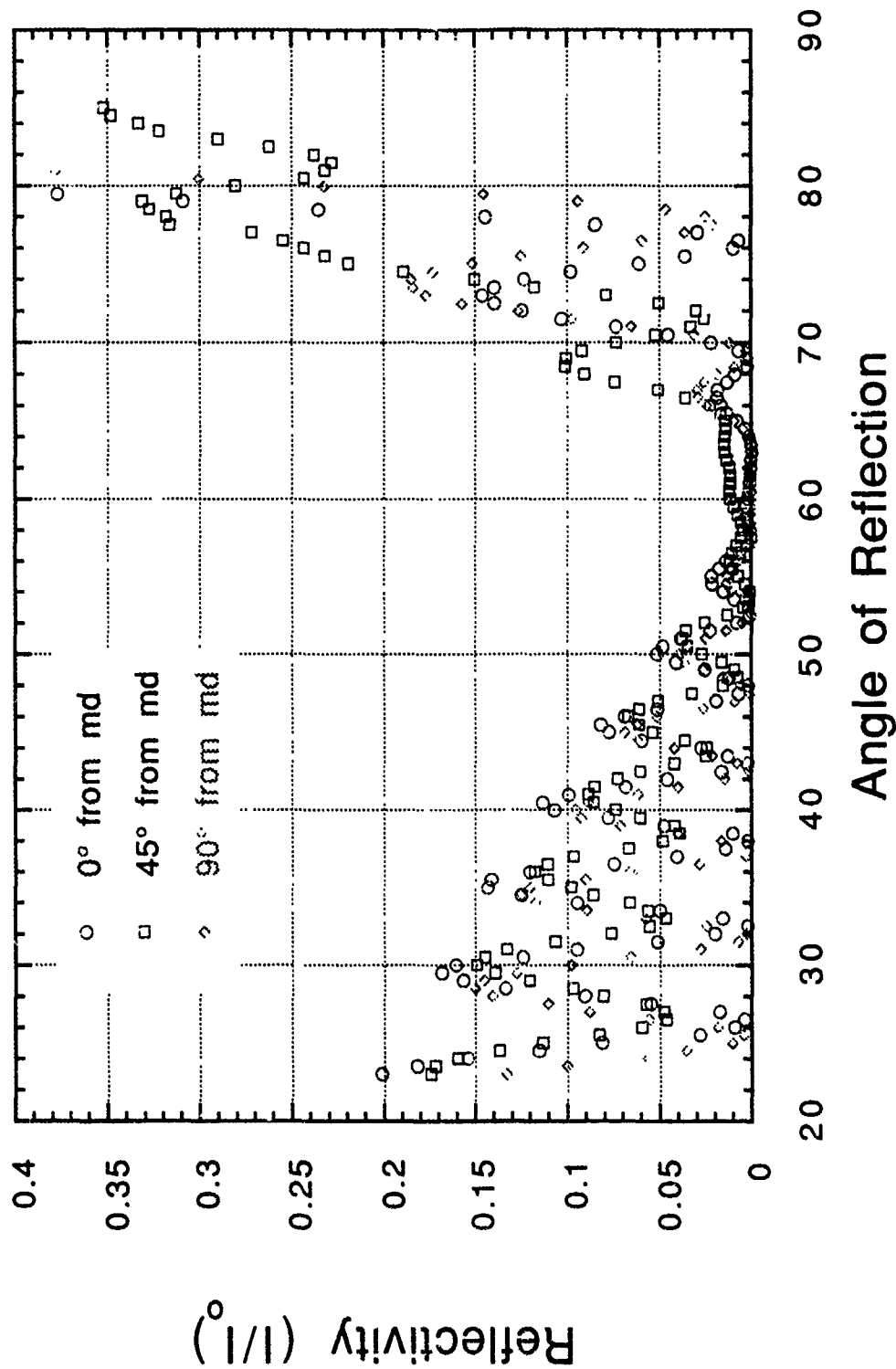
Plot 12

I vs Θ : isotropic Mylar



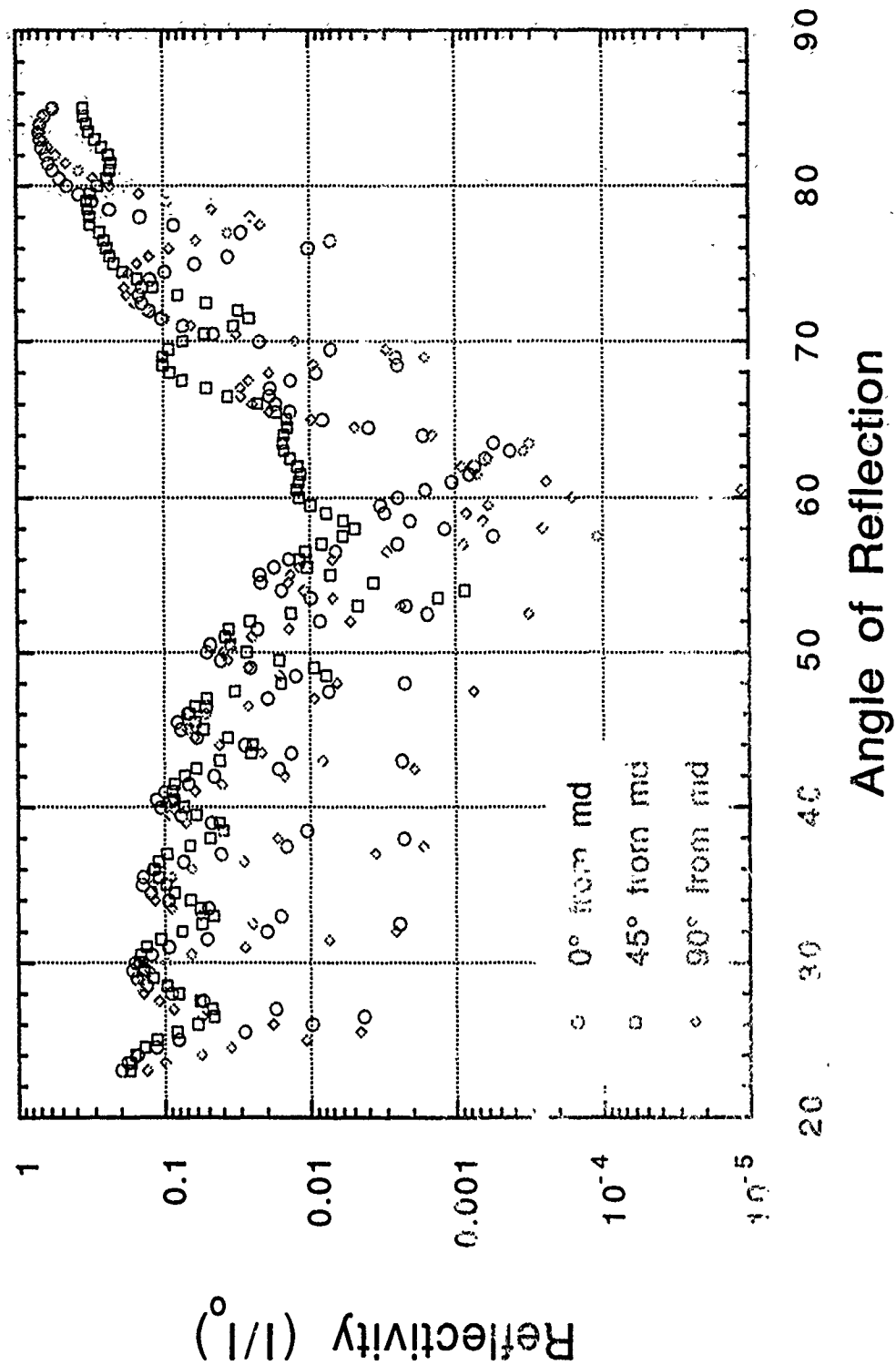
Plot 13

I vs. Θ : anisotropic Mylar



Plot 14

I vs. Θ : anisotropic Mylar



MATHEMATICAL MODELING OF CARBON/CARBON PROCESSING

Jennifer Reid

ABSTRACT

An integral part of expert model development for intelligent processing of composite materials is process modeling. The objective of my summer research was to determine the state-of-the-art development in process modeling of the processing of carbon-carbon composites. With knowledge about the current trends in process modeling and carbon-carbon processing, I was able to evaluate existing models according to their accuracy, capabilities, ease of use, transportability, and the types of principles used to base the models on.

INTRODUCTION

Carbon-carbon composites, which are a class of high performance materials discovered by accident in the late 1950's, can withstand temperatures as high as 6000 F and retain dimensional stability, initial properties, high strength and stiffness, and chemical inertness (9). Due to the exceptional properties exhibited by carbon-carbon composites, the material has many high temperature applications which include aeronautical, space, missile, and industrial.

Carbon-carbon composites consist of a carbon fiber reinforced by carbon matrix. The carbon-carbon composite, prior to a long and costly processing schedule, starts as a composite of carbon precursor fibers in a matrix of thermosetting plastic such as phenolic. The procedure used to obtain the composite prior to processing includes the lay-up and autoclave cure of a liquid impregnated preform. After the initial composite is cured, it is processed until the desired carbon-carbon composite results. The processing consists of the carbonization cycle and the graphitization cycle. During the carbonization cycle, which may last several hours or days, the cured initial composite is subjected to pyrolysis. Pyrolysis is the thermal process which chemically changes organic precursor materials into carbon. Pyrolysis generally occurs in an inert environment between 800 C and 2800 C. The graphitization cycle which follows the carbonization cycle converts the carbon fibers to graphite fibers, which are of higher modulus, between the temperatures of 1900 C and 2800 C. The carbonization and graphitization cycles are the most critical processing steps in the carbon-carbon composite manufacturing scheme, thus three to eight cycles of each might be necessary to achieve the desired high temperature composite.

Since the processing of carbon-carbon composites is both costly and time consuming, a mathematical model which predicts the thermochemical response of a composite during the processing procedure would be extremely valuable. Process modeling predicts the chemical and physical phenomena changes during a process of interest. These phenomena include heat transfer, resin flow, void nucleation and growth, kinetic reaction, and residual stress formation. The trial and error procedure which generally

occurs in the processing of carbon-carbon composites is reduced with process modeling, allowing the effects of the process parameters on the process outcome to be studied.

DISCUSSION OF PROBLEM

After the need for a thermochemical model for the carbonization cycle of carbon-carbon composite processing was defined, an extensive literature search was performed to identify existing models. The search included technical reports, journals, magazines, and an on-line computer literature search in several engineering and scientific data bases. Only the pertinent references are presented in the Appendix, and a more comprehensive list of carbon/carbon literature is available from the author. One theoretical model and associated computer code was located. The Process Environment Model (PEM) was developed to simulate thermochemical and pore pressure responses in two-dimensional carbon-carbon materials during processing.

The Process Environment Model, which was obtained from Dr. W. C. Loomis of Scientific Applications International Corporation, consists of both a one-dimensional and a two-dimensional response coded in FORTRAN. The one-dimensional model (PEM 1-D) is a limited and modified version of the two-dimensional model. PEM 1-D was the one studied because it was the most transportable, not requiring specialized software and a large computer. The one-dimensional version operates on the IBM PC-AT computer or compatible, thus the memory capacity requires that the model be divided into two separate codes. The TMODEL predicts the temperature response of

the composite during processing, while the PMODEL predicts the pore pressure response. Both codes utilize the finite element method of calculation. The PMODEL contains large arrays which require a higher version of FORTRAN and a larger memory which was not available in the laboratory. The PMODEL was then installed and is now operational on a IBM PS-2 Model 80 with FORTRAN 5.1 at the University of Dayton. Only the TMODEL was critically examined.

The TMODEL considers thermochemistry and heat transfer to predict the temperature and degree of carbonization as a function of position and time. The TMODEL consists of seventeen subroutines which are listed in Table 1 (1).

TABLE 1. LIST OF TMODEL SUBROUTINES

TMODEL	Main Program for Thermal Eqn.
DATAIN	Read and Write Input Data
GETXY	Generates Linear Element Integrals
SUMMARY	Write Response Summary
INTERP	Linear Interpolation
BETAI	Calculates Extent of Pyrolysis
ELMINP	Define Element Properties
LINELM	Define Non-Varying Element Eqn. Terms
SYSMAT	Build System Matrix
SYSVEC	Build System Vector
GETELD	Define Variable Element Eqn. Terms
CRANK	Time Step Calculation
INVERT	Generate Matrix Inverse
RAPSON	Newton-Raphson Iteration
GETGE	Build Iteration Matrix
PRINT	Write Real Number Matrix
IPRINT	Write Integer Number Matrix

The input to the TMODEL consists of the control information

which defines the type of problem to be analyzed, the material properties data, kinetic thermochemical data, boundary conditions and associated time tables, and the geometry of the model which locates the model coordinates either automatically or manually. A sample input with variable definitions is given in the Appendix (1).

With a general understanding of the TMODEL code, parametric studies were made in order to evaluate the thermochemical model performance in terms of its accuracy, capabilities, and ease of use. Parameters of particular interest include the number of nodes, the time step, the kinetic parameters, composite properties, and the cure cycle. Only one parameter was changed at a time while holding the others constant. Table 2 contains the particular parameter changes which were made to the input file.

TABLE 2. Parametric Studies

CASE 2	Time Step Interval Decreased
CASE 3	Length of Sample Increased
CASE 4	Length of Sample Decreased
CASE 8	Increase the Step Time to Final Carbonization Time (Base Case)
CASE 9	Change the Cure Cycle
CASE 10	Change the Cure Cycle
CASE 11	Increase the Activation Energy
CASE 12	Increase the Composite Density
CASE 13	Increase the Resin Density
CASE 14	Increase the Reaction Rate Coefficient
CASE 15	Decrease the Convection Coefficient
CASE 16	Add the Radiation Coefficient

RESULTS

The results gained from the parametric studies of the TMODEL computer code are summarized in this section. The input file of the base case, CASE 8, is listed in the Appendix as the sample input.

When the time step was decreased, CASE 2, no significant change in the output of the TMODEL was observed. This indicates that a wide range of time steps can be used without affecting the accuracy of the numerical solution.

When the length of the specimen was changed, CASE 3 and CASE 4, the output reflected the increase or decrease in the resistance to heat transfer respectively, as can be seen in Figure 33 where the temperatures and degree of carbonization (DOC) profiles of the center of the specimen are depicted. Before the onset of the reaction the temperature of the thick specimen lags behind that of the thin specimen. After the exothermic reactions start, the thin specimen temperature is lower than that of the thick specimen. This is due to the lower resistance of the thin specimen allowing easy transfer of the generated heat to the surroundings.

The cure cycle which the carbon-carbon composite is subjected to during the carbonization was altered to analyze the effects of changing the boundary conditions on the temperature and DOC profiles. The new cure cycle which is contained in the input file for CASE 9 is listed in Table 3.

TABLE 3. CASE 9 CURE CYCLE

TIME(hrs)	TEMPERATURE(C)
0.0	25.0
4.0	200.0
30.0	500.0
60.0	590.0
120.0	750.0
140.0	999.0
160.0	1010.0

The temperature and degree of cure at the middle layer for both CASE 8, the base case, and CASE 9 can be observed in Figure 1. The temperature profile for both CASE 8 and CASE 9 reflect what was programmed for the cure cycle. Although both cases reach complete carbonization, CASE 9 completely carbonizes in less time due to the increased heating rate. Table 4 lists the cure cycle which was used in CASE 10.

TABLE 4

TIME(hrs)	TEMPERATURE(C)
0.0	25.0
4.0	150.0
15.0	230.0
30.0	300.0
60.0	450.0
80.0	750.0
120.0	900.0

As noted for CASE 9, the change in the cure cycle, CASE 10, causes a change in the output of the thermochemical model. Figure 2

illustrates the difference between CASE 8 and CASE 10 for the temperature and DOC at the middle layer. The temperature cycles of both CASE 8 and CASE 10 represent the respective cure cycles. Both cases become completely carbonized, but CASE 10 reaches carbonization sooner due to its cure cycle.

The activation energies of the carbonization kinetics are increased by a factor of ten in CASE 11. The change in the output can be found in Figure 3. As expected, the change in temperature as a function of time for CASE 11 and CASE 8 are almost identical due to their similar cure cycle. Although the change in the activation energy does not affect the heating rate of this particular case of the model, it has a significant effect on the degree of carbonization. The energy barrier for the reaction was increased due to the increased activation energy, resulting in the slowing down of the reaction to a level where the heat generated by the reaction is negligible compared to the heat supplied by the autoclave.

The composite density as a function of temperature was changed in CASE 12. The effect of the new density on the temperature and degree of carbonization profiles are shown in Figure 4. The temperature profile of CASE 12 indicates that the composite does not reach the expected temperature during the cure cycle. When the density increases, the resistance to heat transfer through the material increases, causing the temperature of the denser material (CASE 12) to lag behind that of the material for CASE 10. In addition, the degree of carbonization for CASE 12 is less than that of CASE 8. The specimen does not completely carbonize since the composite can not reach the desired temperatures specified by the

cure cycle.

Figure 5 depicts the changes in the output due a change in the resin density as a function of temperature, CASE 13, and CASE 8. The changes in the resin density cause an exotherm to occur as can be seen in the temperature profile. The exotherm can be explained by the energy equation written for one-dimensional unsteady systems with constant properties.

$$\frac{\partial T}{\partial t} = \frac{k}{\rho C_p} \frac{\partial^2 T}{\partial x^2} = \alpha \frac{\partial^2 T}{\partial x^2}$$

where: k is the thermal conductivity in W/m C
 C_p is the heat capacity in J/kg C
 ρ is the density in kg/m³
 α is the thermal diffusivity in m²/sec

As the density increases, the value of alpha decreases and the value of the change in temperature as a function of time must increase to maintain equation continuity. Thus, the exotherm occurs as the temperature as a function of time suddenly increase in order to compensate for the increased resin density.

The reaction rate coefficients are changed in CASE 14. Figure 6 represents the change in the thermochemical output due to the new reaction rate coefficients. The temperature profiles for CASE 8 and CASE 14 appear to be almost identical, indicating that the change in the reaction rate coefficient for this case did not affect the output temperature of the model (for this particular case, the change in the heat generated by the reaction was negligible compared to the heat supplied by the autoclave).

The degree of carbonization profile for CASE 14 differs from that of the base case. The reaction rate coefficient is directly proportional to the speed of the kinetic reaction, as stated in the rate law.

for an nth order reaction, for example: $-r_A = k(T) C_A^n$

where: $-r_A$ is the reaction rate in mole/m³.hr
 $k(T)$ is the reaction rate constant in (m³/mole)ⁿ⁻¹ / hr
 C_A is the concentration in mole/m³
 n is the reaction order

Therefore complete carbonization of the composite specimen is achieved faster with an increased reaction rate coefficient and kinetic reaction rate.

The convection coefficient was decreased in CASE 15, and the results are presented in Figure 7. The decrease in the convection coefficient caused the temperature of the middle layer to remain almost completely fixed at the initial value. In turn, only a very small amount of carbonization took place. The low value of the convection coefficient did not allow heat to enter into the system, rendering the boundary an insulated boundary.

CONCLUSIONS

A one-dimensional thermochemical model developed by Science Applications International Corporation was installed on a Zenith 248 and is operational. This model predicts the temperature and degree of carbonization as a function of position and time. The pressure model is

also operational.

Parametric studies were performed on the TMODEL in order to establish the accuracy, capabilities, and ease of use of the mathematical model. The model predictions were proven reasonable in terms of the trends in the output in comparison to those predicted by the laws of physics. When a parameter such as the reaction rate coefficient is changed, for example, the kinetic reaction rate should also increase, thus altering the degree of carbonization. The parametric studies which were performed on the TMODEL indicated that the model was in a general agreement with the laws of physics. The capabilities of the TMODEL were investigated by altering the number of nodes, the time step, kinetic parameters, composite properties, and the cure cycle. The code remained operational with all the parameter changes, indicating the wide variety of parameters ranges exhibited by the TMODEL.

The TMODEL is relatively easy to use especially with the input file help generated by this study. The general usefulness of the code is however hindered by the fact that many input parameters such as the kinetic data and the weight losses as a function of temperature are usually not available a priori. This model can however be used as part of an on-line expert model control of the process, and the unavailability of kinetic and weight losses data can be overcome by in-situ measurement of these parameters as the process is taking place.

RECOMMENDATIONS

The pressure model which is now operational must be tested for accuracy and ease of use. The same approach used with the TMODEL can be implemented with the PMODEL.

Actual experimental data of temperature, pressure, and degree of carbonization as a function of time must be used for verification of the accuracy of this model.

A model that can predict weight losses based on kinetic rates must be developed and coupled with the TMODEL. Such a model would eliminate the limitation of the TMODEL which currently requires weight loss data as input parameters.

Kinetic models for other carbon/phenolic systems could be developed and easily incorporated in the TMODEL increasing the applicability of this model to predicting the carbonization of other systems.

REFERENCES

1. Loomis, W. C., F. I. Clayton, and C. N. Heightland, Process Chemistry for 2-D Carbon-Carbon Fabrication for Solid Rocket Motor Technology, Final Report, SAIC-88/1966, November 1988.
2. Loomis, W. C., F.K. Tso, R. B. Dirling Jr., and C.N. Heightland, Analytical Processing for Improved Composites, 5th Interim

Report, AFWAL-TR-4147, Volume II, Analytical Modeling,
October 1980.

3. Brown S. C., et al., Process Science for 2-D Carbon-Carbon Exit Cones, AFWAL-TR-86-4028, July 1986.
4. Pardee, W. J., M. R. Mitchell, A. Gupta, F. Montgomery, and J. Sheehan, Effect of Carbonization Kinetics on In-Process Mechanical Properties, Rockwell International, SC5516.FR.
5. ASM International, Engineered Materials Handbook: Composites, Volume I, Metals Park, Ohio.
6. Schmidt, D. L., Carbon-Carbon Composites for High Temperature Applications, ASD/AFWAL/Materials Laboratory, Wright Patterson AFB, Ohio.
7. Fitzner, E., "The Future of Carbon-Carbon Composites" in Carbon, Volume 5, Number 2, Great Britain, 1987.
8. Pollock, P. B., Structure and Behavior in 2-D Carbon-Carbon, 1987 JANAF Rocket Nozzle Subcommittee Meeting, October 1987.
9. Kliner, K., Carbon-Carbon Composites, Report, University of Dayton, November 1989.
10. Loomis, W. C., F. K. Tso, R. B. Dirling, Jr., and C. N. Heightland, Analytical Processing for Improved Composites, AFWAL-TR-81-4082, Volume I, August 1981.

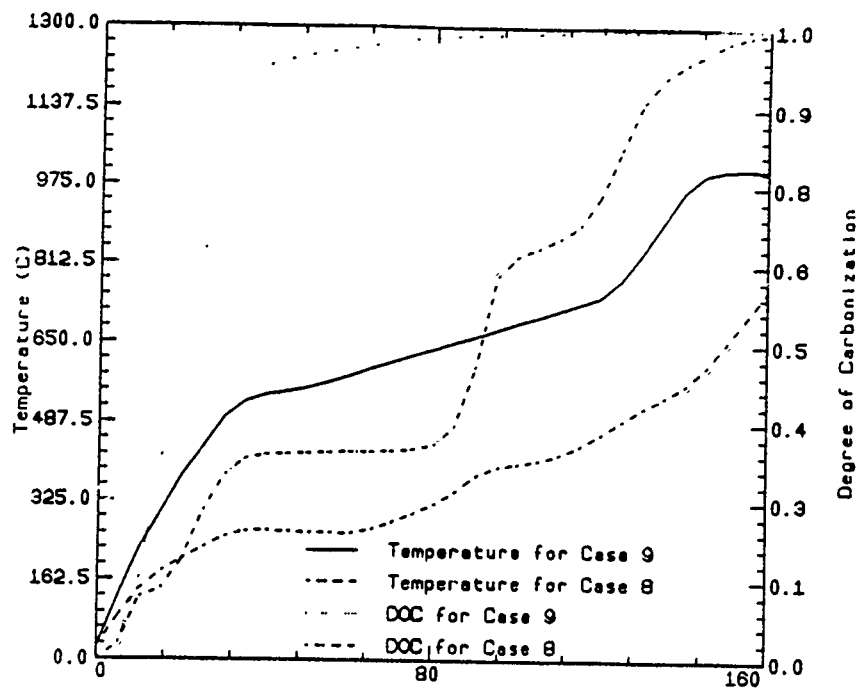


Figure 1. Effect of Changing the Cure Cycle

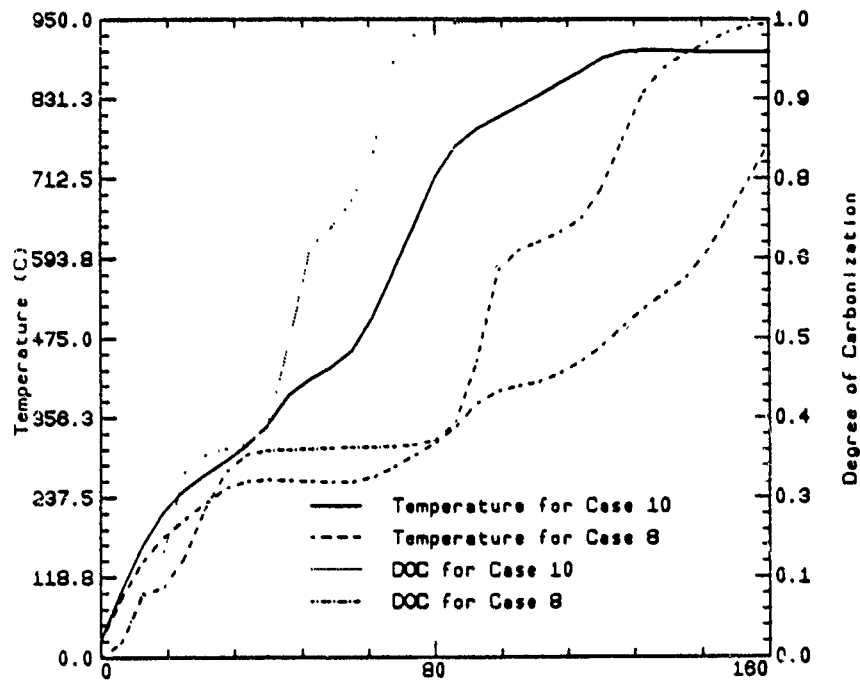


Figure 2. Effect of Changing the Cure Cycle

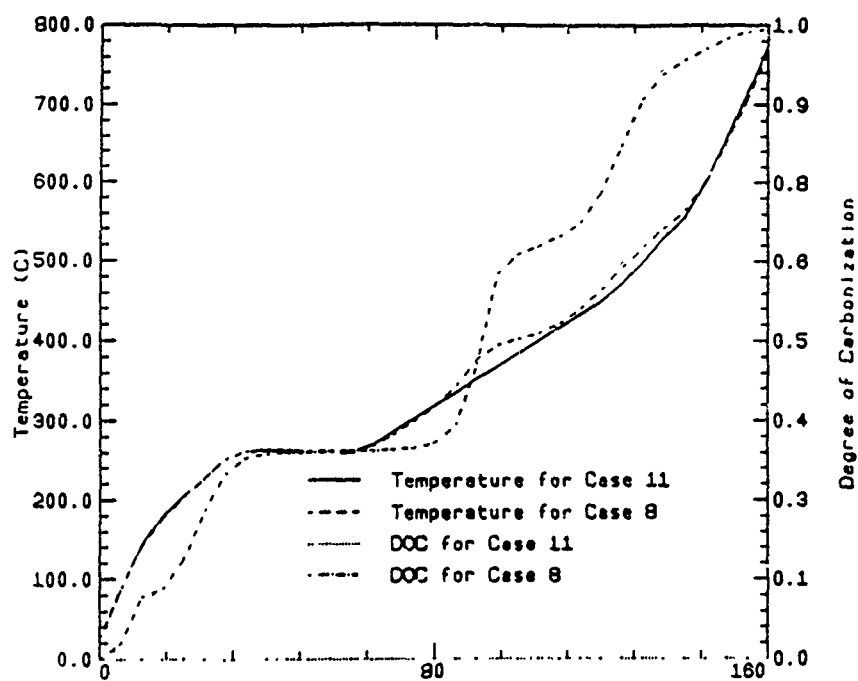


Figure 3. Effect of Increasing the Activation Energy

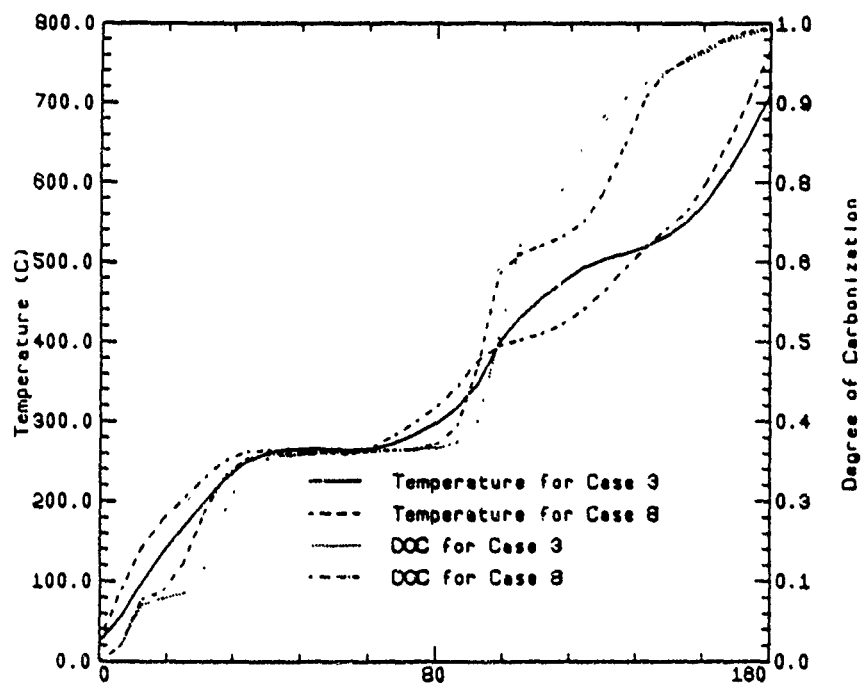


Figure 33. Effect of Changing the Sample Thickness

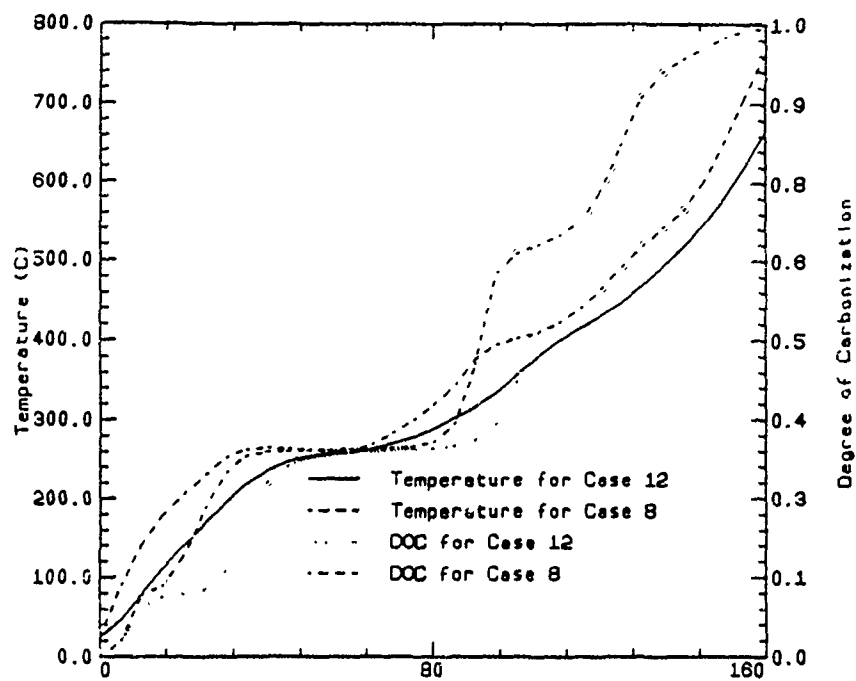


Figure 4. Effect of Increasing Composite Density

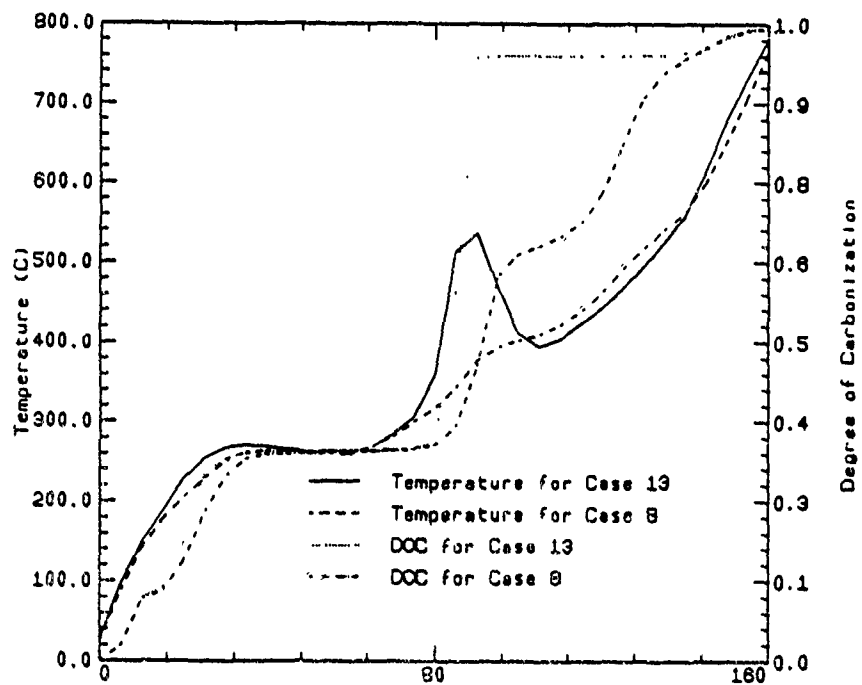


Figure 5. Effect of Increasing Resin Density

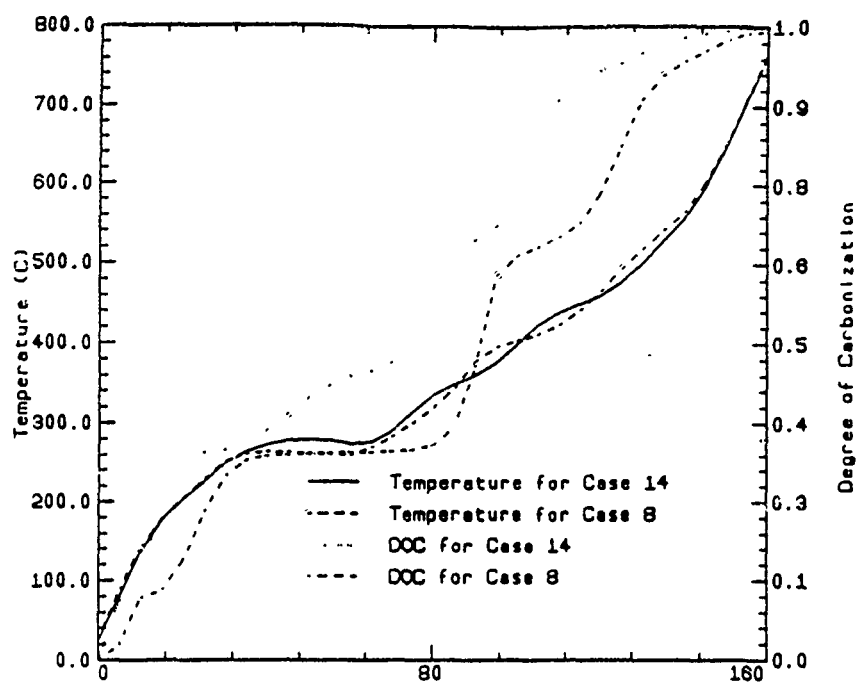


Figure 6. Effect of Increasing the Reaction Rate Constant

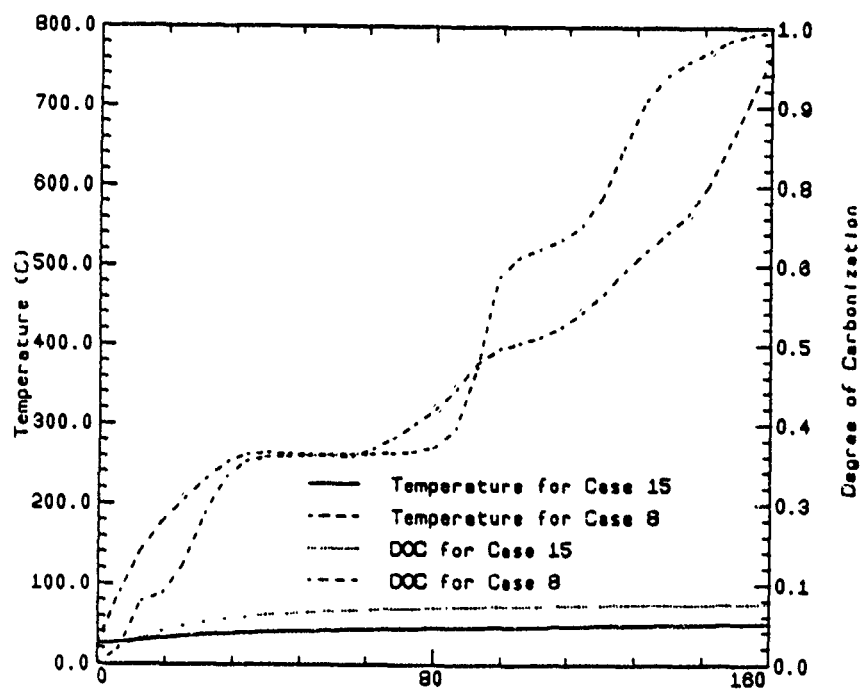


Figure 7. Effect of Decreasing the Heat Transfer Coefficient

APPENDIX

CASE 8 SAMPLE INPUT

TITLE SUPERTEMP PROCESS CYCLE
 #ELEMENTS, #BETA vs To PAIRS, #REACTIONS, NEW OR CONTINUE
 RUN, NODAL SPACE
 35, 21, 4, 0, 0.005
 START TIME(hrs), STOP TIME(hrs), TIME STEP INTERVAL(hrs), MIN
 PRINT INTERVAL(hrs), MAX TIME STEPS(hrs), MAX ITERATIONS
 0.0, 10.0, 0.5, 1, 350, 10
 CONSTANT INITIAL TEMPERATURE(C)
 25.0
 NODE POINT LOCATIONS X1 - Xn(cm)
 0.0, 3.75
 INITIAL RESIN VOLUME FRACTION
 0.29
 NUMBER OF CARBONIZATION TEMPS OF PROPERTY
 8
 CARBONIZATION TEMPERATURES(C)
 100., 200., 300., 400., 500., 600., 700., 800.
 COMPOSITE THERMAL CONDUCTIVITY(kcal/hr-cm-C)
 0.01368, 0.01467, 0.012816, 0.011740, 0.007715, 0.009932,
 0.010552, 0.012366
 NUMBER OF CARBONIZATION TEMPS OF PROPERTY
 8
 CARBONIZATION TEMPERATURES(C)
 100., 200., 300., 400., 500., 600., 700., 800.
 COMPOSITE SPECIFIC HEAT(kcal/g-C)
 .2733E-03, .347241E-03, .368981E-03, .384988E-03,
 .390841E-03, .411266E-03, .424047E-03, .439576E-03
 NUMBER OF CARBONIZATION TEMPS OF PROPERTY
 5
 CARBONIZATION TEMPERATURES(C)
 300., 400., 600., 700., 800.
 COMPOSITE DENSITY(g/cm3)
 1.3912, 1.2913, 1.1743, 1.1549, 1.1475
 NUMBER OF CARBONIZATION TEMPS OF PROPERTY
 7
 CARBONIZATION TEMPERATURES(C)
 200., 300., 400., 500., 600., 700., 800.
 RESIN DENSITY(g/cm3)
 1.27, 1.25, 1.34, 1.27, 1.43, 1.63, 1.72
 FOR FIRST REACTION: REACTION RATE COEFF(hr-1), ACTIVATION
 ENERGY(kcal/gmole), KINETIC EXPONENT, INITIAL RESIN WT
 FRACTION, FINAL RESIN WT FRACTION, MIN HEAT OF RXN ALLOWED,

MAX HEAT OF RXN ALLOWED 0.4257E08, 12.0, 2.6, 0.05, 0.0,
 0.0270, 0.0270
 POLYNOMIAL COEFFS FOR HEAT OF REACTION(kcal/g resin)
 0.0, 0.0, 0.0, 0.0
 FOR SECOND REACTION: REACTION RATE COEFF, ACTIVATION ENERGY,
 KINETIC EXPONENT, INITIAL RESIN WT FRACTION, MIN HEAT OF RXN
 ALLOWED, MAX HEAT OF RXN ALLOWED 0.4181E13, 29.9, 2.2,
 0.1125, 0.0, 0.0625, 0.0625
 POLYNOMIAL COEFFS FOR HEATS OF REACTION(kcal/g resin)
 0.0, 0.0, 0.0, 0.0
 FOR THIRD REACTION: REACTION RATE COEFF, ACTIVATION ENERGY,
 KINETIC EXPONENT, INITIAL RESIN WT FRACTION, FINAL RESIN WT
 FRACTION, MIN HEAT OF RXN ALLOWED, MAX HEAT OF RXN ALLOWED
 0.4702E19, 54.4, 2.5, 0.4362, 0.2737, 0.6222, 0.6222
 POLYNOMIAL COEFFS FOR HEATS OF REACTION(kcal/g resin)
 0.0, 0.0, 0.0, 0.0
 FOR FOURTH REACTION: REACTION RATE COEFF, ACTIVATION ENERGY,
 KINETIC EXPONENT, INITIAL RESIN WT FRACTION, FINAL RESIN WT
 FRACTION, MIN HEAT OF RXN ALLOWED, MAX HEAT OF RXN ALLOWED
 0.9923E21, 68.9, 4.5, 0.4013, 0.237, 0.6222, 0.6222
 POLYNOMIAL COEFFS FOR HEATS OF REACTION(kcal/g resin)
 0.0, 0.0, 0.0, 0.0
 LIST OF BETA VALUES
 0.00029, 0.00428, 0.01063, 0.05641, 0.07831, 0.09988,
 0.12353, 0.24767, 0.28781, 0.33152, 0.36665, 0.5382, 0.60041,
 0.6922, 0.80458, 0.85347, 0.89787, 0.9278, 0.95701, 0.97742,
 0.99173
 LIST OF To VALUES
 25.0, 50.0, 65.0, 115.0, 145.0, 210.0, 240.0, 290.0, 310.0,
 360.0, 385.0, 430.0, 450.0, 485.0, 535.0, 555.0, 580.0,
 605.0, 645.0, 700.0, 800.0
 LIST OF DEL To/ DEL BETA VALUES
 86402.4, 3850.1, 1938.1, 1038.24, 1785.67, 2815.38, 784.77,
 382.33, 613.62, 1331.0, 481.64, 251.43, 387.58, 508.02,
 357.22, 451.0, 662.3, 984.44, 1807.03, 3876.44, 12897.61
 NUMBER OF ORDERED(TIME/TEMP) PAIRS FOR BOUNDARY CONDITIONS,
 CONVECTION COEFF(kcal/hr-cm2-C), RADIATION COEFF(kcal/hr-cm2-
 C) 7, 8.0E-4, 0.0
 TIME, TEMPERATURE FOR BOUNDARY CONDITIONS
 0.0, 25.0, 4.0, 132.0, 30.0, 260.0, 60.0, 260.0, 120.0,
 454.6, 140.0, 565.7, 160.0, 793.5

AUTOMATED INSPECTION CODE GENERATION

Steven M. Ruegsegger
Graduate Student - Electrical Engineering

ABSTRACT

This paper describes the summer research done at the Manufacturing Research group at Wright Patterson Air Force Base (WL/MLIM) through a contract from Research & Development Laboratories (RDL). After some background knowledge is introduced, a discussion of the process leading to code-generated automated inspection plans is developed. An example of a manually produced inspection process plan for an actual part, started from an engineering drawing through to CMM code using the Rapid Design System (RDS), is given for clarity.

INTRODUCTION

Problem Statement

The goal of this work is the development of an automated inspection planner for a CIM system. The inspection planner takes the part model -- consisting of both geometry and tolerance information -- and produces an inspection plan, which gives detailed instructions as to how to inspect a manufactured part to determine whether it is within tolerance. An inspection plan, in this project, will consist of both printed instructions to be manually performed by an inspector, as well as code that be executed by automated inspection equipment, in this case a computer controlled coordinate measuring machine (CMM).

Motivation

It is quite time-consuming for the inspector to understand a part from a drawing and determine how to inspect it, program the CMM, and then evaluate the results. This overhead is particularly critical when it is multiplied by many small lot sizes. An automated

inspection system, that can understand the part model, can reduce these tedious, time consuming, and expensive jobs.

Researching Solution

This project is part of a larger effort to develop a Rapid Design System (RDS), whose objective is to reduce the time from design to manufacturing and inspection. The RDS is being developed with the cooperation of the 4950th Test Wing, which is a design and manufacturing organization which specializes in custom modification of aircraft production of replacement parts which are not available from the manufacturer. The typical scenario is the production of very small lots (1 to 100) of steel or aluminum machined parts required for the further operation of the Air Force's aircraft. For this type of application, turnaround time -- time it takes from design to acceptable inspection results of a workpiece -- is more important than minimizing total machine time for manufacture and inspection. Here we expect the current typical lead time of 44 days to be cut in half.

DISCUSSION OF THE PROBLEM

Background

I was hired by RDL to be contracted to Wright Patterson Air Force Base (WPAFB). However, the job was already waiting here for me. This project is part of a very large project -- the RDS -- involving WPAFB, Case Western Reserve University (CWRU), and a few other universities. As a recent graduate of CWRU, I am working closely with a professor (my advisor) and his group of students, referred to by *we*, on our own module of the RDS: the Inspection Planning and Evaluation module (IPE).

The RDS is an intelligent software system which couples feature-based-design with Artificial Intelligence to produce design guidelines and suggestions, and automated manufacturing and inspection process plans. It does this by providing an intelligent CAD interface which enables the designer to get his design into the computer faster and more

accurately than current systems allow, and also integrated tools to check a design for manufacturability and inspectibility, then upon acceptance of the design, produce the optimal process plans.

The CAD system developed in the RDS uses the concept of "**feature-based design**" and thus called the Feature Based Design Environment (FBDE), where a part is described in terms of design features, which represent higher-level concepts than the geometric primitives used in traditional systems, *e.g.* lines, arcs, points. One class of these design features, form features, includes slots, ribs, bosses, through-holes, blind-holes, and pockets. These correspond to specific geometric configurations on the surface of the part-model. Other types of design features include form feature modifiers and tolerance features. There are also manufacturing features and inspection features that are used only in their respective modules.

The inspection process is driven by tolerances specified by the designer which are input through the FBDE. Tolerances are structured according to the ANSI Y14.5 standard called **Geometric Dimensioning and Tolerancing (GD&T)** which consists of two tolerance types: coordinate and geometric tolerances. A coordinate tolerance is defined as the traditional plus/minus tolerance that requires the distance between two points to be evaluated. Geometric tolerances are defined to describe allowable variations in the pose of a geometric feature or can be used to specify allowable variation in the intrinsic properties of a feature, *e.g.* the flatness of a surface.

GD&T is driven by datums. Datums can be either surfaces or curves, including the axis of a hole or boss. These datums can then be combined to form a datum reference frame (DRF). These DRFs connote the functionality of the part by establishing the primary, secondary, and tertiary datums which orient the part in the position was made to perform in and allows a more accurate measurement and evaluation.

My responsibilities in the development of the RDS reside within the **Inspection Planning and Evaluation (IPE)** module. Inspection consists of performing manual

and/or automated operations to evaluate the specified tolerances. The complete sequence of such operations, when executed, will result in the statement that a part meets or fails to meet all specified tolerances. Additional information, such as which tolerances were not met, etc., are anticipated to also be provided.

The IPE is also designed to operate with features. These inspection features are created from the design features which are the inputs to the IPE from the FBDE. Then these inspection features can be expanded, organized, reviewed, sequenced, and manipulated to produce the actual inspection code. For this project, the inspection code is a software language that will drive a Coordinate Measuring Machine (CMM) to properly inspect and evaluate the workpiece.

A CMM is basically a gantry robot with a two degree-of-freedom wrist (see Figure 1). This mechanical arm can position a touch probe to determine the location of a point (very accurately) on the inspected part's surface. The evaluation of the inspection data is the most involved part of the inspection process. CMM's are designed for gathering very accurate data points; however, then mathematical algorithms must be designed to properly evaluate the data points into the geometric primitives and then be able to determine if that primitive is within the allowable tolerance.

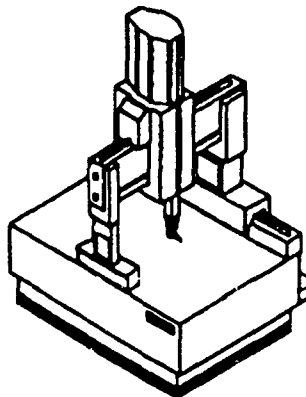


Figure 1 - LK Tool G-80C CMM

The 4950th purchased a CMM from a company called LK Tool. It is controlled from a dedicated 386 clone and driven by a software language called CMES (Co-ordinate Measuring Software) which was written by LK Tool specifically for their machines.

Job Expectations

This summer I was responsible for the output of the IPE: the inspection process's actual instruction set. While the other engineers were working on the procedures from the input to output, I was assigned to work from the output to the input (and sort of meet them in the middle). The output is the CMES language which will drive the CMM and also evaluate its results. This is a brand-new language for everyone and had to be learned from scratch.

There were two basic expectations of this summer job. First was to understand CMES enough to be able to just operate the CMM to probe and return the data points. This is described in a training manual we received from the 4950th. The QC (Quality Control) engineers at the 4950th had gotten this much understanding from a training course they took. The next step was to be able to perform and evaluate the GD&T tolerance callouts (inspection features) using the CMES language. This was a clean-room engineering task, since there is no work ever done on it.

Work Performed This Summer

To start the summer, there was a lot of **background research** to do. I had to get up to speed on such topics as process planning, probe path planning, and sequencing. I received plenty of journal papers and articles about these topics. I also read about and experimented with the RDS system itself. Part of this included studying the programming language LISP. Common LISP an object-oriented software language that the RDS is written in and used in conjunction with a parametric engineering design system called the Concept Modeller™ from Wisdom Systems.

As mentioned, CMES is the software language that controls the LK Tool CMM at the 4950th Test Wing. The industry standard for CMM controlling languages is DMIS. CMES is LK Tool's own derivation of DMIS; therefore, the only documentation is from LK Tool. I received a reference manual and a beginners training manual from the 4950th that they received from LK Tool. These were very limited sources, but did allow me to achieve the first expectation of being able to perform the basic probe and retrieve operations of the CMM. I was able to work regularly with the engineers in QC. This is as far as they had gotten also, due to time and resource limitations. Therefore, in order for them to inspect a machined part, they would simply probe the part's various surfaces, retrieve the raw data, and then use their hand-held calculators to do their own evaluation -- a very tedious, time-consuming, and expensive process. This would take several weeks just to be able to start one lot, and then have to inspect and evaluate it by hand.

The second expectation was fulfilled by going one step beyond this basic knowledge of CMES. It required a lot more reading and studying to learn how to structure CMES to also do the evaluation that was defined by the GD&T callout.

A close study of GD&T had to be done this summer in order to understand what each callout means from an inspection point of view. There are many books on GD&T, but the matter of using a CMM to calculate the callout is almost nonexistent.

There are two basic concepts of GD&T that had to be fully understood in order to be implemented into CMES. The first, and most extensive, was the DRF (Datum Reference Frame) which provides primary, secondary, and tertiary datums used to remove all degrees of freedom from the workpiece and create a stable, repeatable inspection fixture for the workpiece. However, for a automated inspection, more than just this is required. An origin of a coordinate system is required as well as approachable faces for the probe. I wrote an algorithm (drf2crf) which determines whether a designer has input a valid DRF, and then convert it into a proper CRF -- coordinate reference frame -- for the CMM to use.

The other concept of GD&T was the individual callouts. Within GD&T, some callouts refer to surfaces and others to features. We had to distinguish between these and also make sure that this was properly done within the FBDE tolerance features and then converted properly to the inspection features.

The next, and most important, step this summer was to combine the studies discussed above: process planning, CMES language, and GD&T tolerancing and evaluation. This required a full understanding of what it means to inspect each GD&T tolerance with respect to each feature/surface, and then writing the CMES instructions to do so in a efficient and accurate manner. This will be coded into LISP that writes the CMES code, *e.g.* code-generating code. To get to this step, I developed "generic" CMES code that, for each tolerance callout/feature combination, would act as a macro. This CMES "macro" is created with slots (variables) for the exact coordinal reference values (x,y,z) to be substituted for when the macro is instantiated. All these generic codes, once instantiated, are concatenated to help make up the automated inspection process.

An example of the generic CMES code is given in Figure 2. Note that the CMES code is in capital letters, while the LISP variables are in lower case. I could not include all the features/tolerance combinations (too long), but did include one sample and a list of all the features and tolerances that were included.

As a note here, only geometric GD&T callout tolerances have been implemented into the generic CMES code. The next step is to allow the coordinal tolerances (\pm tolerances) to be coded. Here, each design feature will have all the inspection knowledge on how to inspect itself and the designer/inspector operating the RDS will chose which tolerances he wants measured. Some of this I studied during the summer and I even got a working model for one feature, but it will take time and inspection experience to go through all the design features. Also the implementation of coordinal tolerance in the FBDE has been moved to a low priority.

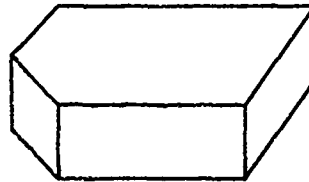
edge-cut

MC move course

PP probe w/ compensation

UG use group n1 + n2 more

AN angularity comparison



```
∠      ! assume datum is already created
      #MC\X1\Y1\Z1
      #PP,axis\Z0      \SP,1
      { ... n total times (2 lines): 3 ≤ n ≤ 40 ... }
      UG,1,n\AX,4,n
      AN,4,datum\angle,tol(.5)
```

```
⊥      ! assume datum is already created
      #MC\X1\Y1\Z1
      #PP,axis\Z0      \SP,1
      { ... n total times (2 lines): 3 ≤ n ≤ 40 ... }
      UG,1,n\AX,4,n
      N1,4,datum\Aol(.5)
      ! or AN,4,datum/A90,tol(.5)
```

Figure 2 - Sample of Generic CMES code

RESULTS

Throughout the summer, the best way to study and then test my material was to create actual inspection plans on actual parts. Many of these inspection plans were manually generated in order to discover the process behind the plan and then be able to program that into the inspection module of the RDS. I have enclosed one of these manual study examples to demonstrate the RDS from the FBDE through to the actual CMES code.

The input into the RDS is given through the FBDE. The user starts with an engineering drawing, an actual part model, or just an idea. Here is the engineering drawing of the part model that is toleranced using GD&T callouts as well as coordinal tolerances (see Figure 3).

Technical drawing of a mechanical part, showing front and side views with dimensions and feature callouts.

Front View Dimensions:

- Overall width: 1.750"
- Overall height: 4.58"
- Top flange width: 1.250"
- Top flange thickness: .500"
- Top flange hole diameter: $2 \times \varnothing .250^{+0.003}_{-0.000}$
- Top flange hole position: $\varnothing .003 (M)$ A B C
- Top flange hole position: -D-
- Top flange hole position: -B-
- Top flange hole position: -C-
- Top flange hole position: .500"
- Top flange hole position: .500"
- Top flange hole position: .500"
- Top flange hole position: .250"
- Top flange hole position: 2.375"
- Top flange hole position: .625"
- Top flange hole position: .875"
- Top flange hole position: 1.750"

Side View Dimensions:

- Overall width: 3/4"
- Overall height: 4.58"

Feature Callouts:

- Top flange hole: $2 \times \varnothing .250^{+0.003}_{-0.000}$
- Top flange hole: $\varnothing .003 (M)$ A B C
- Top flange hole: -D-
- Top flange hole: -B-
- Top flange hole: -C-
- Top flange hole: .500"
- Top flange hole: .500"
- Top flange hole: .500"
- Top flange hole: .250"
- Top flange hole: 2.375"
- Top flange hole: .625"
- Top flange hole: .875"
- Top flange hole: 1.750"
- Top flange hole: $\varnothing .430^{+0.004}_{-0.000}$
- Top flange hole: $\varnothing .005 (M)$ A D (M) E (M)
- Top flange hole: 4X $\varnothing .206^{+0.003}_{-0.000}$
- Top flange hole: $\varnothing .005 (M)$ A D (M) E (M)
- Top flange hole: -E-
- Top flange hole: -A-

Appendix A shows a snapshot of the FBDE module of the RDS after the part model has been loaded into it. Notice how it is a prismatic starting block with various features added to it. All the GD&T tolerance features are added as well; however, the FBDE does not display them as yet.



After the user has finished creating the part and tolerancing it within the FBDE, he initiates the inspection planner by mouse clicking on the appropriate button. Within the IPE (see Figure 4), the tolerance features are copied into inspection features and then manipulated into Inspection Plan Requests (IPRs). Each IPR contains the information needed to inspect that feature/surface with that tolerance. Below is a text representation of a partial list (due to space constraints) of the IPRs for this part.

```
((IPR-NUMBER 8) (IPR-TYPE "POSITION TOLERANCE")
  (CONST-ON-ORIENTATION-LIST NIL)
    #<MEASUREMENT-RQST-A 171276665>
    #<Standard-Class MEASUREMENT-RQST-A 202616446>:
    :INSTANCE allocation:
    GEOM-TO-INSPECT      268436024
    SURFACE-TYPE         CYLINDRICAL
    NUMBER-OF-PTS        (3 2)
    CONSTRAINT-ON-PTS-LIST (CO-PLANAR)

    #<EVALUATION-RQST 171301026>
    #<Standard-Class EVALUATION-RQST 202617046>:
    :INSTANCE allocation:
    EVALUATION-FUNC      FIT-PLANE
    EVALUATION-CONST     "unbound"
    TOL-ZONE-SHAPE       CYL-TOL-ZONE
    TOL-VALUE            0.003

    #<COMPARISON-RQST 171301356>
    #<Standard-Class COMPARISON-RQST 220717426>:
    :INSTANCE allocation:
    COMPARISON-TYPE      NIL
```

1)

Next, the IPRs are combined, sorted, and ordered into what is called an Inspection Plan Fragment (IPF). Now the series of IPFs are in the format to be converted directly to the inspection plan output -- CMM code. This is performed through the generic CMES code macros. Each IPF contains a pointer to the proper macro and then initiates it, which calculates the proper values to plug into the variable slots. Then a CMES header and footer are added and the output is sent to a window for the inspector to check over and edit any specific needs that he has. Below (shown in a two column format) is the entire CMES inspection plan for this *simple* part. Many parts take up to 10 pages of CMES code just because they contain many more features and tolerances.

CMES PROGRAM TO INSPECT
HOUSEPLATE /MODEL #1

=====

CREATED FROM THE RAPID DESIGN SYSTEM
DESIGNER: Steve Ruegsegger

DATE: 8 Jul 91

DA

\\\\

MOVE PROBE CLEAR OF OBSTRUCTIONS

\\\\

ENTER TI WHEN READY FOR SETUP

\\\\

ET

Set up ICRF (Initial CRF)

\\

Take three points on the top surface CCW (z^)

AX-Z

Take two points on the left surface F TO E (y^)

N1+,Y,Z

Take one point on the back surface (x^)

N2+,X,Y,Z

PNMD\SA,1

\\\\

MOVE PROBE CLEAR OF OBSTRUCTIONS

\\

ENTER TI TO CONTINUE WITH
AUTOMATED INSPECTION

\\

ET

BEGINNING AUTOMATED
INSPECTION

! Recalculate ICRF using auto-inspection

#MC\Z\1.25\25\#PP,Z\0 \SP,1

#MC,Z\25

#MC\3\375\25\#PP,Z\0 \SP,2

#MC,Z\25

#MC,Y\2.125\#PP,Z\0 \SP,3

#MC,Z\25

#MC,Y\-.25\#MC,Z\-.25\#PP,Y\0 \SP,4

#MC,Y\-.25

#MC,X\1.5\#PP,Y\0 \SP,5

#MC,Y,Z\-.25\25

#MC,X,Y\-.25\1.25\#MC,Z\-.25\#PP,X\0 \SP,6

#MC,X,Z\-.25\1.25

!

UG,1,3\AX-Z

UP,4,5\N1+,Y,Z

UP,6\N2+,X,Y,Z

PNMD\SA,1

PA,LP:

Create DRF#1

! (NOP b/c ICRF=DRF#1)

Inspect DRF#1 D4 features

Inspect H1

#ID,ZM\500\1.250\-.500\250,.003,-.000\003

DB\SP,1 \ save datum D w/ MMC

#MC,Z\25

Inspect H2

#ID,ZM\3.750\1.250\-.500\250,.003,-.000\003

DB\SP,2 \ save datum E w/ MMC

#MC,Z\25

Create DRF#2

! NIL

UP,1\SD,X,Y

Inspect DRF#2 D4 features

Inspect H7

#ID,ZM\1.000\0\-.500\430,.004,-.000\005

#MC,Z\25

Inspect H3

#ID,ZM\500\875\-.500\205,.003,-.000\005

#MC,Z\25

Inspect H4

#ID,ZM\2.875\875\-.500\205,.003,-.000\005

#MC,Z\25

Inspect H5

#ID,ZM\2.875\-.875\-.500\205,.003,-.000\005

#MC,Z\25

Inspect H6

#ID,ZM\500\-.500\-.875\-.500\205,.003,-.000\005

#MC,Z\2.25

ending

DB,clear

RM

Done with inspection

EP\LP:

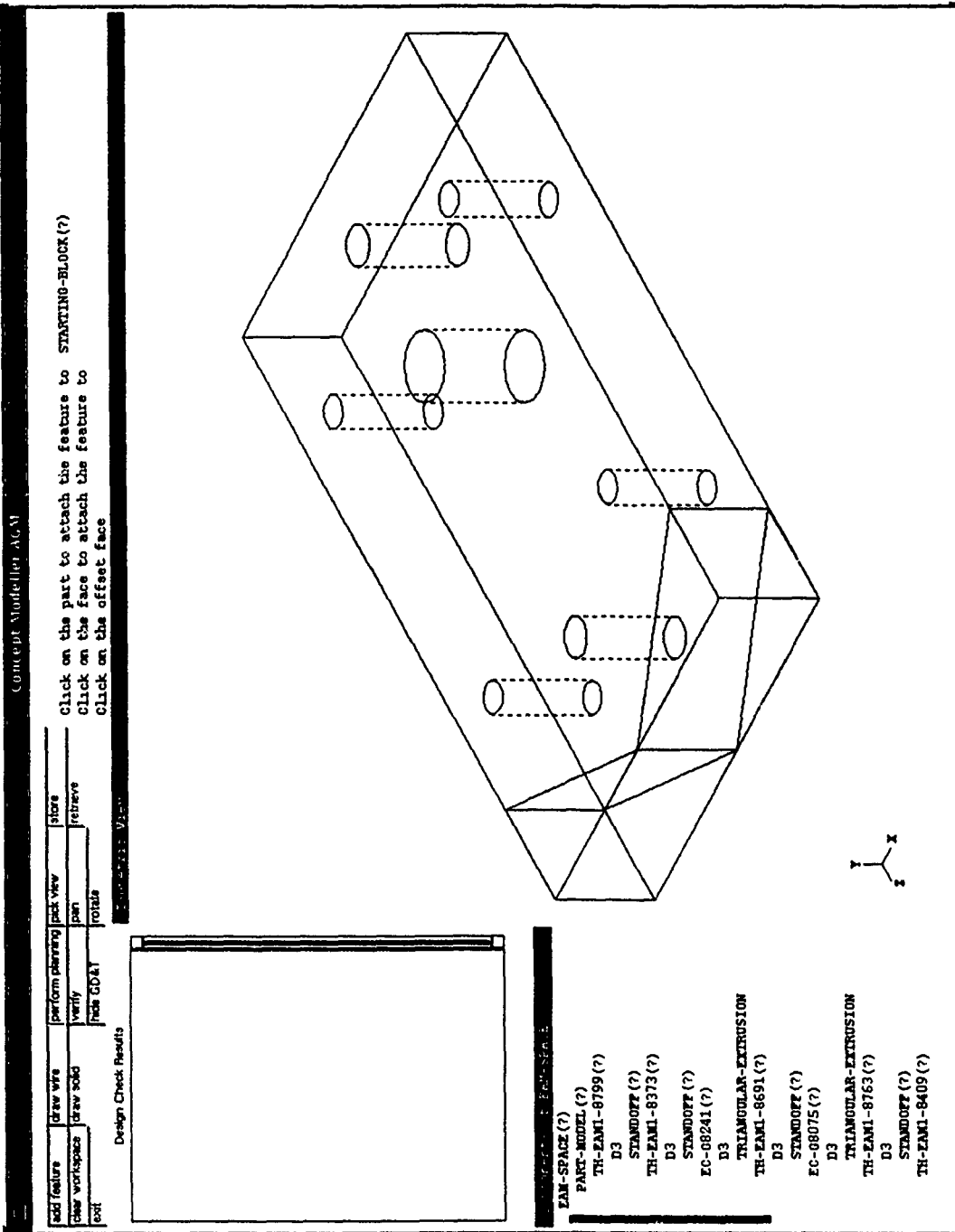
ET

CONCLUSIONS

The goals of this summer were not to come to a definite ending, but as an introduction to my master's thesis with the work here understood to continue for the next year. The real goal of this summer was to help in furthering the development of the

Inspection module so it could to be demonstrated on the October deliverable deadline. I believe that this was a successful summer as we had the fastest growth ever of the IPE.

There is still much to be completed -- mostly programming all these ideas, algorithms, rules, and methods into LISP code for integration into the RDS. A more efficient method of path planning will be needed as well as a program that would display the probe's path overlaid onto the part model. The biggest need, I believe, is inspection experience and trying to capture this experienced knowledge into rules and methods to be programmed into the RDS.



PROCESSING OF MONOLITHIC Ti_5Si_3

**Andrew Thom
Ames Laboratory
Iowa State University
Ames, IA**

August 8, 1991

ABSTRACT

The processing of Ti_5Si_3 to produce a fully dense monolith is investigated. The effect of hot consolidation technique and parameters are considered. The resulting monolith is sufficiently brittle at room temperature to cause pullout of grains during preparation of metallographic samples. Thermal expansion anisotropy, elastic anisotropy, and impurity effects are suspected of causing this behavior.

I. INTRODUCTION

A. Background Information

High temperature structural materials are critical to the continuing development of high performance aerospace systems. Projects such as the DoD Integrated High Performance Turbine Engine Technology initiative (IHPTET) demand materials suitable for service temperatures up to 1600°C. Properties such as specific strength, specific stiffness, creep resistance, oxidation resistance, and microstructural stability are of importance. Refractory intermetallics are among the class of materials currently under investigation.

In the present paper, Ti_5Si_3 has been chosen for investigation because of its promise as a high-temperature structural material. It has a high melting temperature of 2130°C, has adequate oxidation resistance, and has an attractively low density of 4.32 g/cc. Ti_5Si_3 has a hexagonal crystal structure. There are two formula units per unit cell, and has the Mn_5Si_3 structure type with lattice constants $a=7.429\text{\AA}$ and $c=5.1392\text{\AA}$ ¹. The Ti-Si binary phase diagram² shows that Ti_5Si_3 is a congruently melting compound and has a compositional range of 2-4 at% Si.

As an initial step in the investigation of Ti_5Si_3 , the processing necessary to produce a fully dense, monolithic sample is studied. The influence of synthesis and consolidation on the development of a dense monolith is considered. Before outlining the experimental procedure used, a brief review of recent research about processing monolithic Ti_5Si_3 is appropriate.

B. Literature Review

Liu et al.³ investigated the mechanical and microstructural properties of stoichiometric Ti_5Si_3 as well as the effects of microalloying and macroalloying. Synthesis techniques included arc melting/casting and powder metallurgy. Microcavities and microcracks were present in the cast silicide alloys and could not be eliminated by control of the casting process. However, these castings were suitable for post heat treatment for future mechanical testing. The starting powders were semiconductor grade silicon and pure titanium with 400ppm

wt% oxygen, but no information was given on the morphology. The PM technique involved CIPing well mixed powders followed by sintering in vacuum/ Ti_2 . The sintering process resulted in an average volume expansion of 14%. Assuming incorporation of oxygen, the PM technique was not used.

Silicides with 37.5 and 38 at% Si from arc melting/casting were judged to be single phase. Microalloying with boron and carbon at 100-5000 ppm wt% did not decrease microcracking. An alloy of Ti-33Si-4Zr-4Cr (at%) produced almost no cracks. Silicide alloys showed a hardness of 980 dph. Tensile tests in vacuum up to 1000°C showed no indication of ductile behavior. The alloy of Ti-37Si-300ppmC, HIPed at 1600°C for 1 hour at 207 MPa, had a fracture strength of 28 MPa at room temperature and fractured during heating to 800°C. The silicides exhibited mainly transgranular fracture with step and river patterns.

Frommeyer et al.⁴ used PM techniques to successfully produce Ti_5Si_3 . Elemental powders were ground to 10-50 μm under argon and then sealed in a mild steel can. The powders were HIPed at 1250°C for 10 hours at 200 MPa. Fully dense material was produced with only a few micropores on the order of 0.2 μm . The microstructure showed randomly distributed grains on the order of 20-50 μm . Bulk CTE was measured to be $7.1 \times 10^{-6} \text{K}^{-1}$ and $9.1 \times 10^{-6} \text{K}^{-1}$ at 25°C and 1000°C. This is lower than that of $11.0 \times 10^{-6} \text{K}^{-1}$ reported in [5]. Relative to the c-axis and a-axis, CTE was estimated to be $10 \times 10^{-6} \text{K}^{-1}$ and $3 \times 10^{-6} \text{K}^{-1}$, showing a strong anisotropy. The measured bulk modulus was 156 GPa and 140 GPa at 25°C and 1000°C. Modulus along c-axis was estimated at 265 GPa at 600°C, indicating strong anisotropy. Compression testing at 10^{-4} s^{-1} showed no plastic deformation below 1000 °C, but at 1050°C the 0.2% yield stress was 1050 MPa. Fracture toughness determined from four point bending was 2-3 $\text{MPa} \cdot \sqrt{\text{m}}$ from 25°C to 500°C. This low value was attributed to segregation of impurities at the grain boundaries as detected by Auger depth profiling. Segregated oxygen was seen to go into solution with Ti_5Si_3 above 700°C, accounting for an increase in toughness to 7 $\text{MPa} \cdot \sqrt{\text{m}}$.

Reuss et al.⁶ produced Ti_5Si_3 by HIPing elemental powders. The resulting materials were of two differing grain sizes: 10-30 μm and 6 μm . Four point bending tests were used to measure fracture toughness up to 1100°C. Toughness of the fine grain material was better at all temperatures, showing K_{IC} values of about 12 $\text{MPa} \cdot \sqrt{\text{m}}$ at an estimated brittle to ductile transition temperature between 800°C and 900°C.

II. EXPERIMENTAL PROCEDURE

Ti₅Si₃ ingots were produced by argon arc-melting stoichiometric 125 g batches of Ti and Si. These ingots were ground into powder using a WC lined mill. The powders were handled under the inert conditions of an argon filled glovebox. The resulting powders were then sieved. The -325 mesh powders ($\leq 44 \mu\text{m}$) were then characterized by chemical analysis, X-ray diffraction, SEM, and particle size analysis. The chemical analysis is shown in Table 1.

Table 1: Composition of -325 Mesh Powder

<u>Element</u>	<u>Weight %</u>
Si	26.38
Ti	73.02
O	.391
C	.029
N	.077
Nb	.0306
Fe	.0711
Cr	.0035
Al	.0023

The powder is slightly silicon rich, Ti₅Si_{3.08}, and contains an appreciable oxygen impurity of 3900 ppm. XRD verifies the Ti₅Si₃ phase with no unmatched peaks. Particle size analysis shows a bimodal distribution centered at 0.5 μm and 5.0 μm and a specific surface area of 1-2 m²/g. The SEM micrograph in Figure 1 shows the irregular shape of the particles as produced by the grinding operation. The source of the oxygen contamination is currently under investigation. A likely explanation is that the powders were contaminated if exposed to air during handling for chemical analysis.

The powders were then consolidated by hot pressing (HP) and hot isostatic pressing (HIP). Samples were hot pressed in an induction heated vacuum furnace at 1×10^{-1} Torr. Graphite dies were lined with tantalum foil, and temperature was measured with W/Re 5/26 thermocouple. One inch diameter samples of 20 g of powder were pressed at 34 MPa for 4 hours at 1300°C, 1580°C, and 1700°C. Typical time to ramp to soak temperature was 2 hours,

while ramp to room temperature was by natural cooling. Die was loaded with powder in the glovebox and quickly transported to the hot press furnace.

The HIPed powder was loaded into a mild steel can and evacuated to 10^{-3} Torr and degassed at 400°C for 1 hour. The can was then sealed and HIPed in argon at 1200°C for 6 hours at 240 MPa. One hour ramps were used.

III. RESULTS

A. Consolidation

Table 2 shows the effect of temperature on the apparent density and grain size of hot consolidated -325 mesh powder.

Table 2: Density and Grain Size
of Consolidated Samples

<u>Sample</u>	<u>Temp (°C)</u>	<u>ρ/ρ_{th} (%)</u>	<u>Grain size (μm)</u>
9	1200 ^a	—	5-20
8	1300 ^b	97.1	15-40
2	1580 ^b	97.9	200-500
3	1700 ^b	>99	250-750
	^a HIPed	^b Hot pressed	

Even at the low homologous temperatures, very high densities have been achieved. Extensive grain growth has occurred at the elevated temperatures. This may be due to discontinuous grain growth from the bimodal distribution of the starting powders. The density of sample 9 must still be measured, but it is expected to be at least as dense as sample #8. XRD of sample #2 has verified presence of Ti_5Si_3 with no unmatched peaks.

2. Microstructure

Due to the brittle nature of Ti_5Si_3 at lower temperatures, extreme difficulty has been encountered in preparing metallographic specimens for microstructural analysis. Arc melted ingots exhibited substantial cracking.

Upon sectioning of consolidated samples with a diamond wafer blade, extensive pullout of grains has been observed for all samples. Sectioning by EDM also produced extensive grain pullout. After mounting the specimens, more pullout was observed from grinding and rough polishing. Severity of pullout was proportional to grain size with substantial amounts of area pulling out from samples 2 and 3. Figure 2 shows an optical micrograph in polarized light of sample 8. Pullout can be seen from the light areas. Some intragranular porosity can be seen, but the effect of pullout on the apparent porosity is unclear.

Figure 3 shows a backscattered electron image of sample 2. Although the pullout of grains scatters electrons in the valleys and low spots to cause darker regions, the grains appear to possess a uniform phase. Silicon and titanium maps from electron microprobe show a uniform distribution of the Ti and Si in sample 2. Qualitative EDAX of samples 2 and 3 indicate presence of only Ti and Si.

Figure 4 shows the fracture surface of sample 8. Substantial intergranular cracking is evident, possibly occurring during the cooling of the sample from the pressing temperature. Some transgranular cracking can also be seen. A small amount of porosity is present, indicating a very dense sample.

IV. DISCUSSION

Ti₅Si₃ is extremely brittle at low temperatures, as evidenced by the difficulty in the preparing metallographic specimens. There are likely several factors contributing to the pullout of grains during preparation of metallographic samples. Some obvious reasons are thermal expansion anisotropy, elastic anisotropy, and impurities.

In a polycrystalline material with CTE anisotropy, residual stresses will develop during cooling from the process temperature. Additionally, the cooling rate will affect the residual stresses developed as a fast cooling rate will not allow for any relaxation to occur. These residual stresses may cause spontaneous cracking to occur along grain boundaries. Using Frommeyer et al. estimates for CTE, $\Delta \alpha = 7 \times 10^{-6} \text{K}^{-1}$, indicating a substantial anisotropy.

Elastic anisotropy may amplify the stresses due to CTE anisotropy. Tvergaard et al.⁷ have shown that stress singularities occur at triple-grain

junctions which result in a smaller critical grain size for microcracking. Using Frommeyer et al. estimates for elastic modulus, $\Delta E > 100$ GPa, again a substantial anisotropy. The additive stress induced by sectioning, grinding, and polishing may thus be sufficient to cause pullout. It is not surprising that samples 2 and 3 exhibit brittle behavior with grain sizes on the order of hundreds of microns. The brittleness of sample 8 with grain size 15-40 μm and particularly sample 9 with grain size 5-20 μm is disturbing. Frommeyer demonstrated token room temperature toughness for grain size 20-50 μm and Reuss measured the same for grain size 10-30 μm .

Though fracture toughness of the samples from this study has not been measured, it is quite obvious that samples showing such behavior during polishing would give exceedingly low values. Thus elastic and CTE anisotropy are not sufficient alone to explain this behavior. It is possible that the interaction of these anisotropies with impurities are the cause of this phenomenon.

Frommeyer has indicated the deleterious effect of impurities on the fracture toughness of Ti_5Si_3 . K_{IC} values of 2.0-3.5 $\text{MPa}\cdot\sqrt{\text{m}}$ up to 700°C were measured. Oxygen precipitation at the grain boundaries was determined by Auger depth profiling. Above 700°C the oxygen was found to diffuse into Ti_5Si_3 with a corresponding increase in K_{IC} to 7 $\text{MPa}\cdot\sqrt{\text{m}}$ at 1050°C. The solubility of oxygen in Ti_5Si_3 has been shown experimentally by Taylor and Desu⁸ by Auger depth profiling as seen in Figure 5.

While the chemical analysis of the starting powders has shown an oxygen impurity level of 3900 ppm wt%, this measurement is suspect due to the likely exposure of the powder to air during the analysis. The fine powder would have likely absorbed/reacted with atmospheric oxygen. However, this measurement is useful in that the powders prior to hot pressing and HIPing were unavoidably exposed to air prior to consolidation when the loaded die was transferred from the glovebox to the hot press or when the HIP can was transferred from the glovebox to the vacuum sealing equipment. Therefore we may expect that the powders in fact contained an appreciable amount of oxygen prior to consolidation.

V. SUMMARY

Monoliths of near theoretical density have been produced at moderately low homologous temperatures of 0.6-0.7 T_m by hot pressing and HIPing. These monoliths are slightly silicon rich, and have single phase, fine grain sized microstructures. The effect of the anisotropies and impurities on grain pullout has not been established. It seems likely that elastic and CTE anisotropies, perhaps interacting with impurities, are causing the extreme brittle behavior of the monolithic Ti_5Si_3 . A slow controlled cooling rate might allow sufficient relaxation of residual stresses and thus reduce or eliminate the problem of grain pullout. The consolidated samples also require investigation by chemical analysis and Auger depth profiling to quantify the level of oxygen present and qualify the presence of oxygen at the grain boundaries. If reduced cooling rates do not eliminate the problem of grain pullout and grain boundary analysis indicates oxygen segregation, the synthesis and consolidation process must be further investigated. The synthesis technique and consolidation process must then be modified to reduce impurity levels in the powders.



Figure 1: Ti_5Si_3 ingots milled and sieved to -325 mesh.

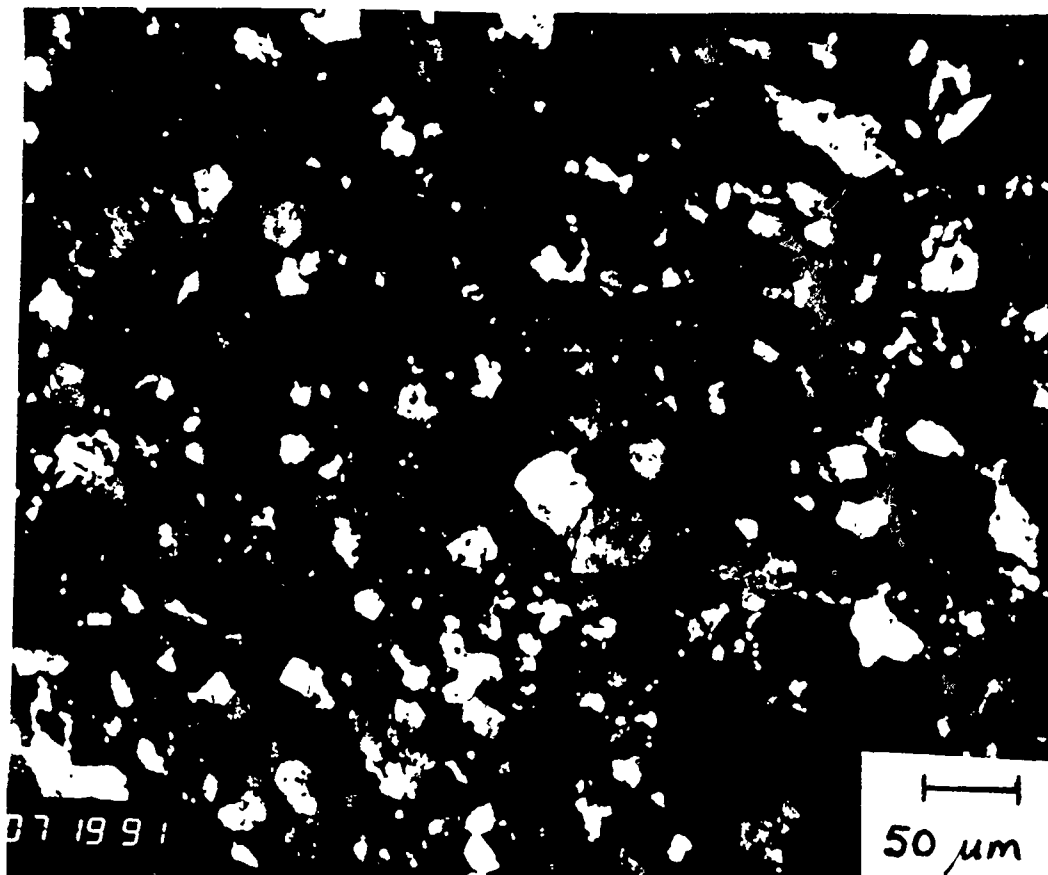


Figure 2: Sample 8, hot pressed at 1300°C
for 4 hours at 5 ksi in vacuum.

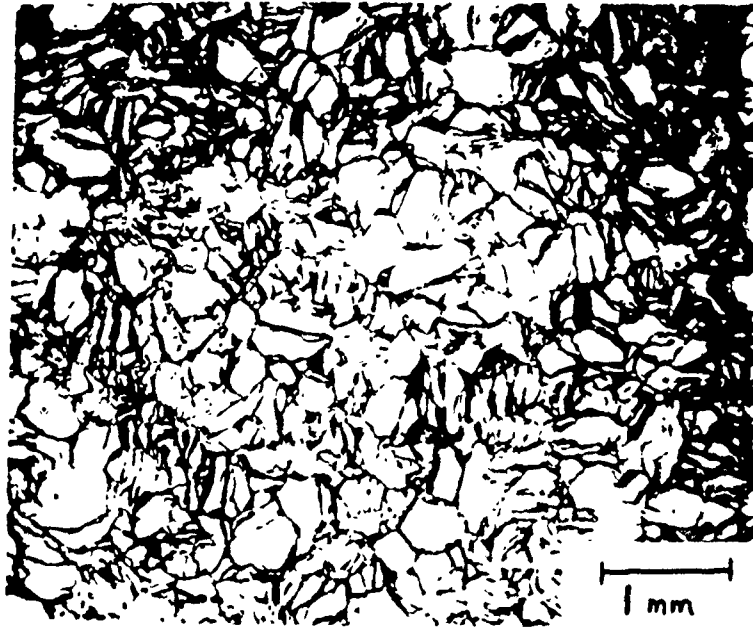


Figure 3: BS image of sample 2, hot pressed at 1580°C for 4 hrs at 5 ksi.

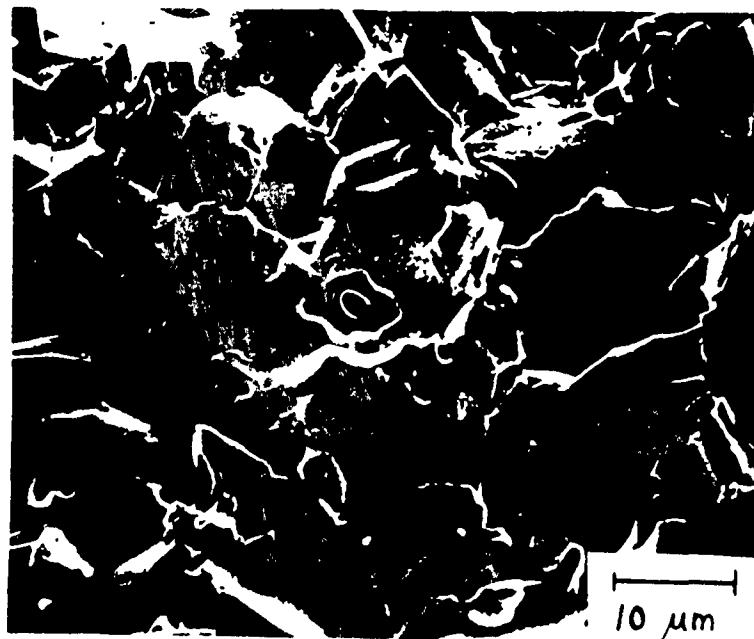


Figure 4: Fracture surface of sample 8, hot pressed at 1300°C for 4 hrs at 5 ksi.

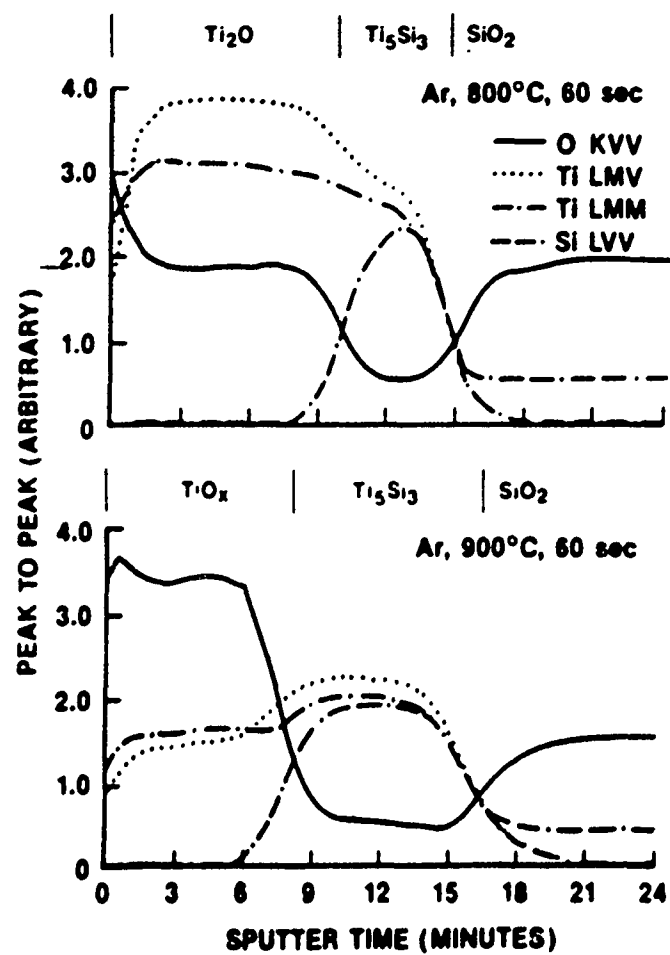


Figure 5: Auger depth profiling of Ti/SiO₂ after annealing in Ar for 60s at 800°C and 900°C.⁸

REFERENCES

1. P. Villars and L. D. Calvert, *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*, American Society of Metals, Metals Park, Ohio 1985.
2. T. B. Massalski (ed.), *Binary Alloy Phase Diagrams*, Vol. 3, American Society of Metals, Metals Park, Ohio (1990) 3370.
3. C. T. Liu, Z. H. Lee, and T. J. Henson, "Initial Development of High-Temperature Titanium Silicide Alloys," ORNL Technical Report No. 6435, 1988.
4. G. Frommeyer, R. Rosenkranz, C. Ludecke, "Microstructure and Properties of the Refractory Intermetallic Ti₅Si₃ Compound and the Unidirectionally Solidified Eutectic Ti-Ti₅Si₃ Alloy," *Z. Metallkde.*, Bd. 81, H. 5, (1990) 307-313.
5. T. Ya. Kosolapova (ed.), *Handbook of High Temperature Compounds: Properties, Production, Applications*, Hemisphere Publishing Corp., New York, New York (1990).
6. S. Reuss and H. Vehoff, "Temperature Dependence of the Fracture Toughness of Single Phase and Two Phase Intermetallics," *Scripta Metallurgica et Materialia*, Vol. 24, (1990) 1021-26.
7. V. Tvergaard and J. W. Hutchinson, "Microcracking in Ceramics Induced by Thermal Expansion or Elastic Anisotropy," *J. Am. Ceram. Soc.*, 71 [3] 157-66 (1988).
8. J. A. Taylor and S. B. Desu, "Interaction of Titanium with Silica After Rapid Thermal Annealing In Argon, Nitrogen, and Oxygen," *J. Am. Ceram. Soc.*, 72 [10] 1947-54 (1989).